

QA
1
C21

CANADIAN JOURNAL OF MATHEMATICS

Journal Canadien de Mathématiques

VOL. I - NO. 3

1949

On the motion of particles in general relativity theory	A. Einstein and L. Infeld	209
Some properties of the eigenfunctions of the Laplace-Operator on Riemannian manifolds	S. Minakshisundaram and Å. Pleijel	242
On the motion of three vortices	J. L. Synge	257
On surface waves	Alexander Weinstein	271
Angular measure and integral curvature	Herbert Busemann	279
The density of reducible integers	S. D. Chowla and John Todd	297
On a theorem of Latimer and MacDuffee	Olga Taussky	300
Congruence relations between the traces of matrix powers	J. S. Frame	303

Published for

THE CANADIAN MATHEMATICAL CONGRESS

by the University of Toronto Press

EDITORIAL BOARD

H. S. M. Coxeter, A. Gauthier, L. Infeld, R. D. James, R. L. Jeffery,
G. de B. Robinson

with the co-operation of

R. Brauer, J. Chapelon, D. B. DeLury, P. Dubreil, I. Halperin,
W. V. D. Hodge, S. MacLane, L. J. Mordell, G. Pall, J. L. Synge,
A. W. Tucker, W. J. Webber

The chief languages of the *Journal* are English and French.

Manuscripts for publication in the *Journal* should be sent to the *Editor-in-Chief*, H. S. M. Coxeter, University of Toronto. Every paper should contain an introduction summarizing the results as far as possible in such a way as to be understood by the non-expert.

All other correspondence should be addressed to the *Managing Editor*, G. de B. Robinson, University of Toronto.

The *Journal* is published quarterly. Subscriptions should be sent to the *Managing Editor*. The price per volume of four numbers is \$6.00. This is reduced to \$3.00 for individuals who are members of the following Societies:

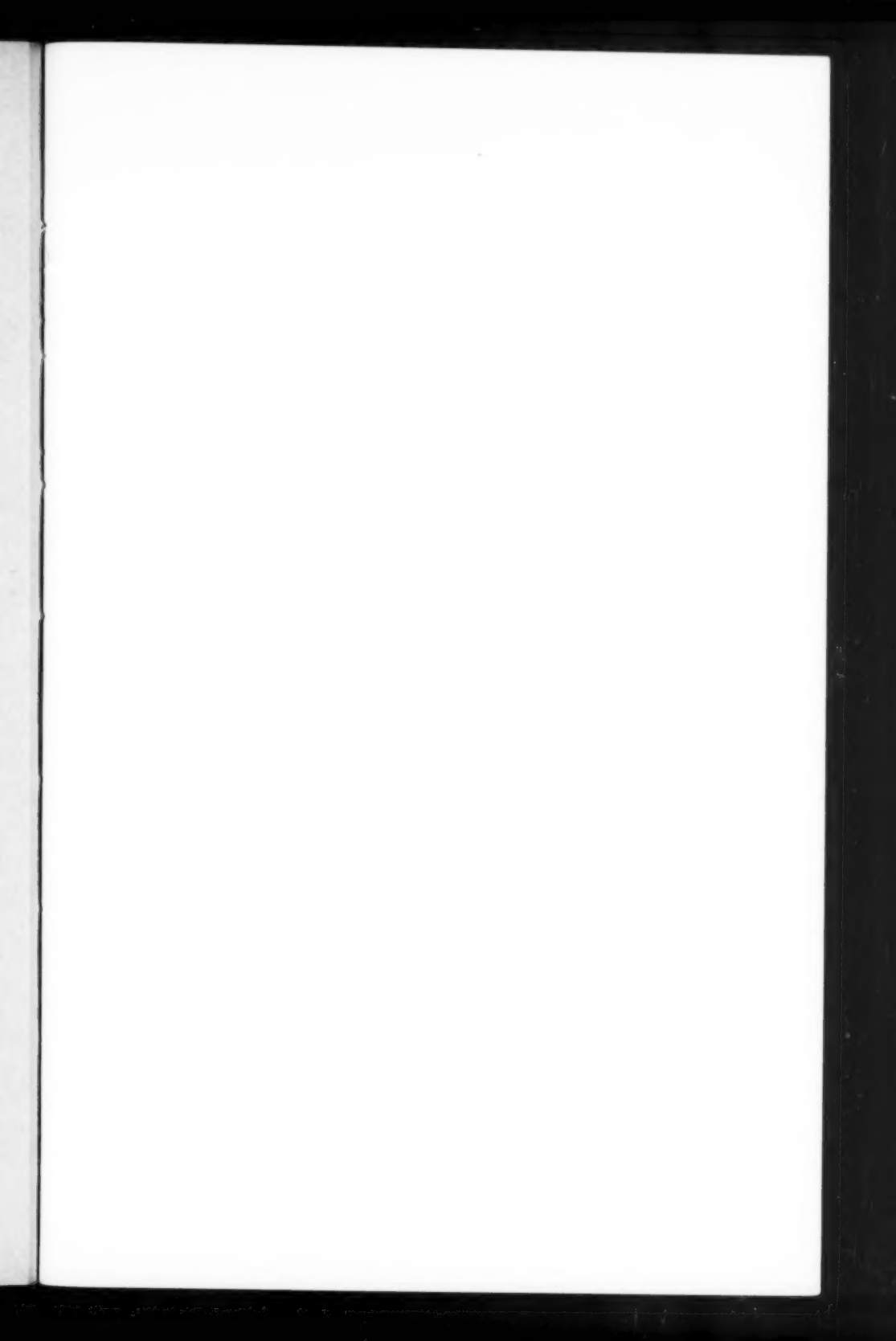
Canadian Mathematical Congress
American Mathematical Society
Mathematical Association of America
London Mathematical Society
Société Mathématique de France

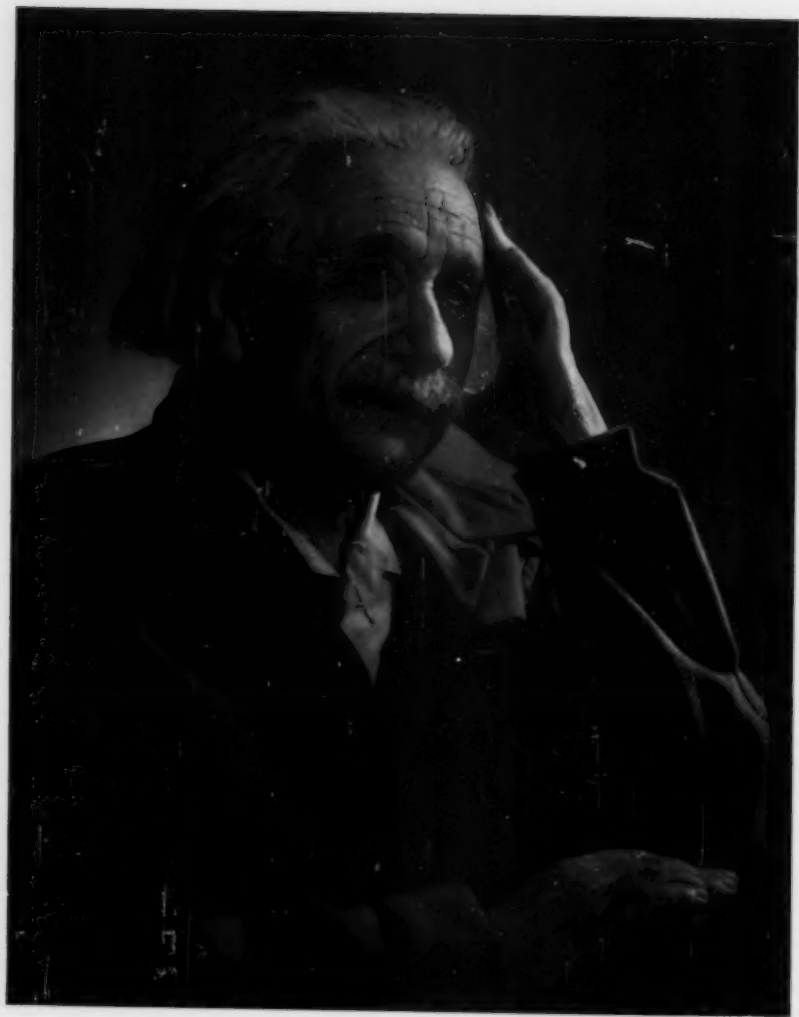
The Canadian Mathematical Congress gratefully acknowledges the assistance of the following towards the cost of publishing this *Journal*:

University of British Columbia	Loyola College
Ecole Polytechnique	University of Manitoba
McGill University	McMaster University
Queen's University	University of Toronto

and

The American Mathematical Society





Kersh—Ottawa

ALBERT EINSTEIN
ANNOS LXX NATUS

ON THE MOTION OF PARTICLES IN GENERAL RELATIVITY THEORY

A. EINSTEIN and L. INFELD

1. Introduction. The gravitational field manifests itself in the motion of bodies. Therefore the problem of determining the motion of such bodies from the field equations alone is of fundamental importance. This problem was solved for the first time some ten years ago and the equations of motion for two particles were then deduced [1]. A more general and simplified version of this problem was given shortly thereafter [2].

Mr. Lewison pointed out to us, that from our approximation procedure, it does not follow that the field equations can be solved up to an arbitrarily high approximation. This is indeed true. We believe that the present work not only removes this difficulty, but that it gives a new and deeper insight into the problem of motion. From the logical point of view the present theory is considerably simpler and clearer than the old one. But as always, we must pay for these logical simplifications by prolonging the chain of technical argument.

The subject matter is presented here from the beginning and the knowledge of previous work is not assumed. To facilitate the reading for those who have studied the previous papers we use here essentially the same notation as before.

Let us start with some general remarks.

All attempts to represent matter by an energy-momentum tensor are unsatisfactory and we wish to free our theory from any particular choice of such a tensor. Therefore we shall deal here only with gravitational equations in empty space, and matter will be represented by singularities of the gravitational field.

In Newtonian mechanics, particles are represented as singularities of a scalar field φ , which satisfies Laplace's equation everywhere outside the singularities. Because the classical equation is linear, the field can be decomposed into partial fields, each part due to a single particle. Each particle *is* in a field due to all other particles. The theory is completed by the equation of motion, that is by putting the acceleration equal to the negative gradient of the field, the proportionality factor being a universal constant. Thus classical physics postulates the equations of motion independently of the field laws. The masses of the sources of the field are assumed to be independent of time. The laws of motion are supposed to be valid in an inertial system. Therefore space-time appears as an independent physical entity. The conceptual weakness of such a space-time background in the classical theory was already recognized by Newton.

Received February 12, 1949.

If we compare this state of affairs with that in general relativity theory, in its original formulation, we see striking similarities and differences. Laplace's equation

$$\Delta\varphi = 0$$

is replaced by the gravitational equation

$$R_{kl} = 0,$$

which, however, unlike the classical equation, satisfies the general relativity principle. The classical principle of inertia becomes in relativity theory the principle of the geodesic line valid for a particle with infinitely small mass. True enough, the difficulty with the inertial system disappears in relativity theory, as does the independent physical reality of space-time. Yet the equations of motion still appear independently of the field equations.

Our aim is to investigate to what extent the field equations *alone* contain the equations of motion of particles; also to develop a method that will allow us to find these equations of motion up to an arbitrary approximation.

Let us start with a simple remark: a *linear* law always means that the motion of singularities is arbitrary. If to a world-line of a singularity with mass m_1 there belongs a field $F_{(1)}$ and if to a world-line of a singularity with mass m_2 there belongs a field $F_{(2)}$, then the superposition of these two fields, that is $F_{(1)} + F_{(2)}$ is also a solution of the linear field equations. In such a solution the same two world-lines would appear together that before appeared singly. Therefore the field with its linear laws cannot imply any interaction between the singularities. Thus only non-linear field equations can provide us with equations of motion because only non-linearity can express the interaction between singularities.

But the argument cannot be reversed. Non-linearity is necessary but not sufficient for the equations of motion to follow from the field equations.

The reason why the gravitational field equations do provide us with equations of motion lies not in their non-linear character alone, but also in the fact that these equations are not independent from each other. Indeed, among the ten components four are free, this being due to the freedom of choice in the co-ordinate system. The ten equations are valid, so to speak, only for six effective functions. They would be inconsistent were it not for the four (Bianchi) identities that they satisfy. This must be so for every relativistic system of equations derived from a variational principle. These identities are (besides the non-linearity) responsible for the *equations of motion being determined by the field equations*.

The ideas leading to the equations of motion are not easy and are mutually interwoven.

One of the essential ideas in this paper is the treatment of gravitational equations by a "new approximation method." In it we treat space and time differently. We regard the changes of the field in time as small compared with those in space. Only then do we arrive at a consistent, manageable set of

equations that can be solved step by step. This idea is not new and was contained in the previous papers.

The other important idea is the deduction of the equations of motion, which are *ordinary* differential equations, from the field equations which are *partial* differential equations. This idea, treated here differently than in the previous papers, leads to the use of surface integrals taken around the singularities of the field. These surface integrals will depend only on the motion of the singularities and not on the shape of the surface.

These and other ideas will be treated in detail in this paper. To make them clear we have decided to delegate all the more tedious calculations to the Appendices. (If we refer, for example, to A.4, this means the Appendix belonging to Sec. 4.) But even so, many straightforward but long calculations had to be omitted. This is especially true for the calculations that lead beyond Newtonian motion. We included here a short section on this subject, just for the sake of completeness. But, as in [1], so here we have to refer those who would like to see the full calculations to the manuscript which is deposited at the Institute for Advanced Study.

Finally we should like to thank Mr. Lewison for his critical study of our previous papers, and Mr. Schild for a careful and critical reading of this manuscript.

2. Notations: the gravitational equations. Since in the greater part of our work, we shall have to separate space and time, our notation will not be the usual four-dimensional one. We make the conventions: Latin indices take the values 1, 2, 3, and they refer to space co-ordinates only. Greek indices refer to both space and time, running over the values 0, 1, 2, 3. Repetition of indices implies summation.

The expression

$$(2.1) \quad g_{\mu\nu|\sigma} \text{ etc. stands for } \frac{\partial g_{\mu\nu}}{\partial x^\sigma} \text{ etc.}$$

At infinity the gravitational field takes the Galilean values $\eta_{\mu\nu}$, that is:

$$(2.2) \quad \eta_{mn} = -\delta_{mn}; \eta_{0n} = 0; \eta_{00} = 1.$$

We write:

$$(2.3) \quad g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}; g^{\mu\nu} = \eta^{\mu\nu} + h^{\mu\nu},$$

where $h_{\mu\nu}$ represents the deviation of space-time from flat space and it is not assumed to be small.

The $h^{\mu\nu}$ can be calculated as functions of $h_{\mu\nu}$ by means of the relation

$$(2.4) \quad g_{\mu\sigma} g^{\sigma\nu} = \delta_\mu^\nu.$$

It turns out to be convenient to replace the h 's by γ 's which are their linear combinations:

$$(2.5) \quad \gamma_{\mu\nu} = h_{\mu\nu} - \frac{1}{2} \eta_{\mu\nu} \eta^{\sigma\rho} h_{\sigma\rho},$$

or more explicitly:

$$(2.6) \quad \gamma_{00} = \frac{1}{2} h_{00} + \frac{1}{2} h_{ss}$$

$$(2.7) \quad \gamma_{0n} = h_{0n}$$

$$(2.8) \quad \gamma_{mn} = h_{mn} - \frac{1}{2} \delta_{mn} h_{ss} + \frac{1}{2} \delta_{mn} h_{00}.$$

This replacement is, of course, not very important but it does simplify the calculations.

Thus we can, throughout, replace the h 's by the γ 's. The equations of the gravitational field for empty space,

$$(2.9) \quad R_{\mu\nu} = 0,$$

can be written (see A.2) in the following way:

$$(2.10) \quad \Phi_{00} + 2\Delta_{00} = 0$$

$$(2.11) \quad \Phi_{0n} + 2\Delta_{0n} = 0$$

$$(2.12) \quad \Phi_{mn} + 2\Delta_{mn} = 0,$$

where:

$$(2.13) \quad \Phi_{00} = -\gamma_{00|ss}$$

$$(2.14) \quad \Phi_{0n} = -\gamma_{0m|ss} + \gamma_{0s|sm}$$

$$(2.15) \quad \Phi_{mn} = -\gamma_{mn|ss} + \gamma_{ms|ns} + \gamma_{ns|ms} - \delta_{mn}\gamma_{rs|rs}$$

and:

$$(2.16) \quad 2\Delta_{00} = \gamma_{sr|sr} + 2\Delta'_{00}$$

$$(2.17) \quad 2\Delta_{0n} = \gamma_{ms|s0} - \gamma_{00|m0} + 2\Delta'_{0n}$$

$$(2.18) \quad 2\Delta_{mn} = -\gamma_{0m|0n} - \gamma_{0n|0m} + 2\delta_{mn}\gamma_{0s|0s} \\ + \gamma_{mn|00} - \delta_{mn}\gamma_{00|00} + 2\Delta'_{mn}.$$

In these formulae, all the linear terms are written out explicitly, while $\Delta'_{\mu\nu}$ stands for all the non-linear terms in the γ 's. The division of the linear expressions into those belonging to $\Phi_{\mu\nu}$ and those belonging to $\Delta_{\mu\nu}$ may seem artificial at this moment. In anticipation of further development, we shall remark here, that, in the actual approximation procedure, by which we shall solve the gravitational equations, these linear terms collected in $\Delta_{\mu\nu}$ will behave like the non-linear terms.

3. Lemma. We mentioned in the introduction that the differential equations of motion will be derived by forming surface integrals. The technique of calculating such surface integrals will reappear many times in this paper and it is based on a lemma to which we shall refer as *the lemma*. Here we shall give its formulation and its proof.

We have a set of functions:

$$(3.1) \quad F_{(as \dots)kl}.$$

It is immaterial whether these functions of x^μ have tensorial character, or not. The bracketed indices are Greek, or Latin, and they will not play any role in our argument. But we do assume that these functions are skew-symmetric in the indices k, l :

$$(3.2) \quad F_{(\dots)kl} = -F_{(\dots)lk}.$$

We now form an integral

$$(3.3) \quad \int_{(S_2)} F_{(\dots)kl} n_k dS$$

over an arbitrary two-dimensional closed surface that does not pass through the singularities of the field. In (3.3)

$$(3.4) \quad n_k = \cos(x^k, \vec{n})$$

are the components of the "normal unit" vector to the surface. The words "normal," and "unit" are used in the conventional sense to designate the corresponding functions of the co-ordinates, which are implied by these terms in Euclidean geometry. They have nothing to do with any particular metric.

Our lemma is:

$$(3.5) \quad \int_{(S_2)} F_{(\dots)kl} n_k dS = 0.$$

We see that the integral (3.3) is certainly independent of the shape of the surface, because

$$(3.6) \quad F_{(\dots)kl} = 0,$$

and because of Green's theorem. We can also write the integral (3.3) in the form

$$(3.7) \quad \int_{(S_2)} \text{curl}_n A dS,$$

where

$$F_{(\dots)23} = A_1; F_{(\dots)31} = A_2; F_{(\dots)12} = A_3.$$

But (3.7) and therefore (3.3) can be changed, by Stokes' theorem, into a line integral over the rim of the surface. If the surface is closed, the rim is of zero length. Therefore, our lemma as expressed by (3.5) is proved.

4. Surface integrals. We treat particles of matter as singularities of the field. Let us assume p particles and the knowledge of their world lines. Thus we denote by

$$(4.1) \quad \xi^s(x^0); s = 1, 2, 3, \dots, p,$$

the world-line of the s th singularity. Here and later, the index written on the top will always label the particular singularity.

The gravitational field, that is the γ 's, will depend on the x^μ 's but also on the ξ 's and their time derivatives. The equations that the γ 's fulfill are

$$(4.2) \quad \Phi_{\mu\nu} + 2\Lambda_{\mu\nu} = 0.$$

At an arbitrary moment x^0 , let us surround the s th singularity, and it alone, by a closed surface. Then:

$$(4.3) \quad \int_s (\Phi_{\mu k} + 2\Lambda_{\mu k}) n_k dS = 0,$$

where the s over the integral indicates here, and later too, that the integral is to be taken on a two-dimensional surface surrounding the s th singularity and it alone.

We shall show that

$$(4.4) \quad \int \Phi_{\mu k} n_k dS = 0.$$

Indeed it follows from the definition (2.14) and (2.15) of $\Phi_{\mu k}$ that it can be written in the following form:

$$(4.5) \quad \Phi_{\mu k} = F_{(\mu)kl|l}$$

$$(4.6) \quad F_{(\mu)kl} = \gamma_{\mu l|k} - \gamma_{\mu k|l} - \delta_{\mu k} \gamma_{lr|r} + \delta_{\mu l} \gamma_{kr|r}.$$

But $F_{(\mu)kl}$ is skew-symmetric in k and l . Therefore (4.4) is fulfilled. From it and from (4.3) we deduce:

$$(4.7) \quad \int 2\Lambda_{\mu k} n_k dS = 0.$$

Also, because of the structure of $\Phi_{\mu k}$ we easily verify:

$$(4.8) \quad \Phi_{\mu n|n} = 0,$$

therefore also:

$$(4.9) \quad \Lambda_{\mu n|n} = 0.$$

Equation (4.9) tells us that no surface integral of the form (4.7) can depend on the shape of the surface. But equation (4.7) tells us more; namely, that such an integral vanishes.

The $4p$ surface integrals in (4.7) can give us no relation between the space co-ordinates of the field, because the surface is entirely arbitrary. They can only give us relations between the co-ordinates of the singularities and their time derivatives. Thus we may have at most $4p$ differential equations. Anticipating the later development, we may remark here that these equations will determine $3p$ functions of time

$$\xi_k^i(x^0),$$

that is, the motion of singularities.

5. The method of approximation. The problem before us is to solve our field equations and to deduce the equations of motion. This we shall do by a new approximation procedure. Let us assume a function $\varphi(x^\mu, \lambda)$ developed into a power series in the parameter λ (for small λ):

$$(5.1) \quad \varphi(x^\mu, \lambda) = \lambda^0 \varphi_0 + \lambda^1 \varphi_1 + \lambda^2 \varphi_2 + \dots = \sum_{l=0}^{\infty} \lambda^l \varphi_l.$$

The indices below indicate the *order* (l in λ^l is always the exponent, not the index).

If the function φ varies quickly in space, but slowly with x^0 , then we are justified in not treating all its derivatives in the same fashion. The derivatives with respect to x^0 will be of a higher order than space derivatives. We can formalize the procedure by introducing an *auxiliary time* τ ,

$$(5.2) \quad \tau = x^0 \lambda,$$

so that derivatives with respect to τ can be treated on the same footing as the space derivatives:

$$(5.3) \quad \varphi|_0 = \frac{\partial \varphi}{\partial x^0} = \frac{\partial \varphi}{\partial \tau} \lambda = \lambda \varphi_{,0}.$$

We conclude: the "stroke differentiation" of a quantity with respect to x^0 , can be replaced by the "comma differentiation" with respect to τ if the power of λ with which this quantity is associated is simultaneously raised by one. To express this explicitly we use numbers under zeros, written after the comma, e.g.:

$$(5.4) \quad \lambda^{2l} \gamma_{mn}|_0 = \lambda^{2l+1} \gamma_{mn,0} \text{ OR: } \lambda^{2l} \gamma_{mn}|_{00} = \lambda^{2l+2} \gamma_{mn,00}.$$

From now on, all differentiations will be with respect to (τ, x^1, x^2, x^3) and they will be denoted by commas:

$$(5.5) \quad \gamma \dots|_s = \gamma \dots, s; \quad \gamma \dots|_0 = \lambda \gamma \dots, 0.$$

Thus we shall develop all functions that appear in the field equations in power series in λ . We start with the γ 's in the following way:

$$(5.6) \quad \begin{cases} \gamma_{00} = \lambda^2 \gamma_{00} + \lambda^4 \gamma_{00} + \lambda^6 \gamma_{00} + \dots \\ \gamma_{0m} = \lambda^3 \gamma_{0m} + \lambda^5 \gamma_{0m} + \dots \\ \gamma_{mn} = \lambda^4 \gamma_{mn} + \lambda^6 \gamma_{mn} + \dots \end{cases}$$

Why do we start with different powers of λ ? This is an assumption, but it can be justified heuristically. Assuming for a moment the usual energy momentum tensor for matter, we have, for a quasi-stationary field, approximately:

$$(5.7) \quad \begin{cases} \Delta \gamma_{00} = -2\rho \\ \Delta \gamma_{0m} = -2\rho \frac{dx^m}{d\tau} \lambda \\ \Delta \gamma_{mn} = -2\rho \frac{dx^m}{d\tau} \frac{dx^n}{d\tau} \lambda^2, \end{cases}$$

therefore

$$(5.8) \quad \gamma_{mn} \sim \lambda \gamma_{0m} \sim \lambda^2 \gamma_{00},$$

and it is pure convention that we start with λ^2 for γ_{00} .

The other question suggested by (5.6) is: why do we omit the odd powers of λ in the developments of γ_{00} , γ_{mn} , and the even powers in γ_{0m} ? Indeed, we could have introduced all powers in (5.6). A more thorough investigation shows that our choice (5.6) means that what we are doing here is similar to the procedure in electro-magnetic theory when we take not the retarded, but the half-retarded plus half-advanced potentials [3].

All the functions that will appear later are gained from the γ 's by summation, multiplication, differentiation. Thus to every component, the following rule applies throughout: Any component having an $\begin{Bmatrix} \text{odd} \\ \text{even} \end{Bmatrix}$ number of zero suffixes will have only $\begin{Bmatrix} \text{odd} \\ \text{even} \end{Bmatrix}$ powers of λ in its expansion.

6. Field equations and the approximation method. We go back to the field equations

$$(6.1) \quad \Phi_{\mu\nu} + 2\Lambda_{\mu\nu} = 0$$

into which we introduce the γ 's in their power-series development. Thus the (00) equation in (6.1) can be written:

$$(6.2) \quad \sum_l \lambda^{2l} (\Phi_{00} + 2\Lambda_{00}) = 0.$$

Now we cut up (6.2), and the other field equations, into equations for each approximation step. We write them down in the following form:

$$(6.3a) \quad \Phi_{00} + 2\Lambda_{00} = 0$$

$$(6.3b) \quad \Phi_{0m} + 2\Lambda_{0m} = 0$$

$$(6.3c) \quad \Phi_{mn} + 2\Lambda_{mn} = 0.$$

Let us analyse more closely the structure of (6.3). Remembering (2.13) to (2.15) we can write more explicitly:

$$(6.4a) \quad \Phi_{00} = -\gamma_{00,rr}$$

$$(6.4b) \quad \Phi_{0m} = -\gamma_{0m,rr} + \gamma_{0r,mr}$$

$$(6.4c) \quad \Phi_{mn} = -\gamma_{mn,rr} + \gamma_{mr, nr} + \gamma_{nr, mr} - \delta_{mn} \gamma_{rs,rs}$$

and:

$$(6.5a) \quad 2\Lambda_{00} = \gamma_{rs,rs} + 2\Lambda'_{00}$$

$$(6.5b) \quad 2\Lambda_{0m} = -\gamma_{00,0m} + \gamma_{mr,0r} + 2\Lambda'_{0m}$$

$$(6.5c) \quad \begin{cases} 2\Lambda_{mn} = -\gamma_{0m,0n} - \gamma_{0n,0m} + 2\delta_{mn}\gamma_{0r,0r} \\ \quad + \gamma_{mn,00} - \delta_{mn}\gamma_{00,00} + 2\Lambda'_{mn} \end{cases}$$

Let us now assume that:

$$(6.6a) \quad \gamma_{00}^{2l-2} \dots \gamma_{00}^{2l-4}$$

$$(6.6b) \quad \gamma_{0m}^{2l-3} \dots \gamma_{0m}^{2l-1}$$

$$(6.6c) \quad \gamma_{mn}^{2l-2} \dots \gamma_{mn}^{2l-1}$$

are all known. Then γ_{00}^{2l-2} can be found from (6.3a). Indeed Λ_{00}^{2l-2} contains only terms already known, since γ_{mn}^{2l-2} is known and Λ'_{00}^{2l-2} is non-linear and can therefore depend only on the known γ 's. The same is true for (6.3b) and (6.3c). The unknown functions are contained in Φ 's; the known functions in the Λ 's. The γ_{00}^{2l-2} , already found from (6.3a), appears as a known function in Λ_{0m}^{2l-1} . Similarly γ_{0m}^{2l-1} found from (6.3b) appears as known in Λ_{mn}^{2l} . Indeed we see now the reasons for our division of linear terms.

Thus our equations (6.3), if solved, will give us

$$(6.7) \quad \gamma_{00}^{2l-2}, \gamma_{0m}^{2l-1}, \gamma_{mn}^{2l}$$

and if such a procedure converges, we can determine the field to any approximation we wish.

The important question to consider is: are the equations (6.3) always solvable?

7. The divergence condition. We go back to our equations (6.3). The first of them, that is

$$(7.1) \quad \Phi_{00}^{2l-2} + 2\Lambda_{00}^{2l-2} = 0$$

is, because of (6.4a) and (6.5a), a Poisson equation, where Λ_{00}^{2l-2} is known. There is no difficulty in integrating this equation and finding γ_{00}^{2l-2} . Next we have (6.3b), and because of (6.4b), we see:

$$(7.2) \quad \Phi_{0m}^{2l-1} = 0.$$

Thus the next three equations can be integrated only if

$$(7.3) \quad \Lambda_{0m}^{2l-1} = 0.$$

But Λ_{0m}^{2l-1} is already known. Therefore we must be sure that our procedure leads us to Λ_{0m}^{2l-1} satisfying (7.3). Similarly the last six equations (6.3c) lead us because of

$$(7.4) \quad \Phi_{mn}^{2l} = 0$$

to the integrability condition:

$$(7.5) \quad \Lambda_{mn}^{2l} = 0.$$

Thus the divergence conditions are satisfied in each approximation step, though not identically. They are satisfied because of the Bianchi identities and because of the previous field equations.

8. The surface condition and the equations of motion. We now approach the most essential part of our argument. We are faced with the task of solving the following system of equations:

$$(8.1a) \quad \Phi_{00} + \frac{2\Lambda_{00}}{2l-2} = 0$$

$$(8.1b) \quad \Phi_{0m} + \frac{2\Lambda_{0m}}{2l-1} = 0$$

$$(8.1c) \quad \Phi_{mn} + \frac{2\Lambda_{mn}}{2l} = 0.$$

We know that because of the Bianchi identities and because (as we assumed) similar equations had been solved in the previous approximations, we have

$$(8.2) \quad \frac{\Lambda_{0m,m}}{2l-1} = 0; \quad \frac{\Lambda_{mn,n}}{2l} = 0.$$

Let us also remember, that there is no difficulty in solving (8.1a) which is a Poisson equation. But what about (8.1b) and (8.1c)?

Before we return to this fundamental question, we wish to discuss the *start* of our approximation procedure which determines the character of our calculations.

In (8.1) we put $l = 2$ and write the first two equations explicitly:

$$(8.3a) \quad \gamma_{00,ss} = 0$$

$$(8.3b) \quad -\gamma_{0m,ss} + \gamma_{0s,ms} = \gamma_{00,1m}.$$

The character of the entire solution will depend on the choice of the harmonic function we take as the solution of (8.3a). As we are interested in solutions representing particles, we shall write:

$$(8.4) \quad \begin{cases} \gamma_{00} = 2\varphi; \quad \varphi = \sum_{s=1}^{\infty} \left\{ -\frac{2m^s}{2} \psi \right\} \\ \psi = [(x^k - \xi^k)(x^k - \bar{\xi}^k)]^{-1/2} = (r)^{-1}. \end{cases}$$

Here r is the "distance" in space of a point from the s th singularity.

We leave it undecided, for the moment, whether $\frac{m^s}{2}$ is a function of time, or a constant. Now we introduce this γ_{00} into (8.3b) and again obtain three equations for the three functions γ_{0m} . But is (8.3b) always solvable? True, the divergence of both sides vanishes. But this is not sufficient. The surface integral of the left-hand side of (8.3b) vanishes, as follows from the lemma. But then the surface integral of the right-hand side of (8.3b) must vanish too.

If we calculate the surface integral around each singularity, we find (see A.4) that it vanishes only if

$$(8.5) \quad \frac{d}{d\tau} \left(\frac{s}{2} \right) = \frac{s}{2} \frac{1}{r} = \frac{s}{2} = 0,$$

that is if the $\frac{s}{2}$'s do not depend on time. This is so, because

$$(8.6) \quad \dot{\psi}_{,0} = -\psi_{,k} \dot{\xi}^k; \quad \left(\dot{\xi}^k = \frac{d\xi^k}{d\tau} \right)$$

and because only expressions proportional to r^{-2} can give a contribution to the surface integral. Thus, going back to (8.4), we have to assume that

$$(8.7) \quad \frac{1}{2} \frac{s}{2}, \frac{2}{2} \frac{s}{2}, \frac{3}{2} \frac{s}{2}, \dots, \frac{p}{2} \frac{s}{2}$$

are constant.

These constants (8.7) can be positive or negative. We shall assume that $\frac{s}{2}$ are positive. Indeed, by taking the first particle and removing all others, we see that $\frac{1}{2} \frac{s}{2}$ is its *gravitational mass*, since for large r the field is that of a particle with gravitational mass $\frac{1}{2} \frac{s}{2}$. This is the same constant of integration that appears in the Schwarzschild solution, since our field for one particle is that of a Schwarzschild singularity when r is large. *Thus we shall have to exclude from our solution negative gravitational masses. But then we must also exclude dipoles and poles of higher order.*

Yet if we try to solve (8.1) we see (the details will be presented later) that we cannot do so without adding certain poles and dipoles to γ_{00} . This we shall have to do, in order to insure the integrability of (8.1) in each approximation. But then the solution of the total field will contain dipoles which are not allowed, since they represent physically meaningless solutions. We shall have to remove them *after* the total field has been calculated. This can be done by *restricting the motion of particles*. That is, the condition that the dipole field vanishes will give us $3p$ ordinary differential equations for the motion of p particles. Thus the motion is undetermined in the approximation procedure. It becomes determined after the approximation procedure is finished and the dipole fields are removed.

In practice, we find solutions both for the field and for the equations of motion only to a certain approximation, say $2n$. We obtain the equations of motion to the $2n$ approximation, by removing all the dipole fields to such an approximation.

Although we have developed our field equations with respect to an arbitrary parameter λ , this λ can be absorbed by the actual equations of motion through the change of scale in $\frac{s}{2}$ and τ , so that λ is absent from the final form of the equations.

We have given a general outline of our treatment. Turning to the details, let us see why (8.1) will not, generally, be integrable. We know, from the contents of Sec. 4, particularly from (4.4) that the surface integrals of the Φ functions vanish. Although this was proved for the total field it is equally true in each approximation step, since the proof made use only of the structure of the Φ 's, which is the same for the total field, as for the field in each approximation. Thus we have:

$$(8.8) \quad \int_{2l-1}^s \Phi_{0r} n_r dS = 0; \quad \int_{2l}^s \Phi_{mr} n_r dS = 0.$$

But then our equations (8.1) can be self-consistent, only if we have:

$$(8.9) \quad \int_{2l-1}^s 2\Lambda_{0r} n_r dS = 0; \quad \int_{2l}^s 2\Lambda_{mr} n_r dS = 0.$$

But the Λ 's in (8.1) are already known; they are functions of the *known* field calculated in the previous approximation steps. Therefore we can calculate the integrals (8.9) and find whether they vanish or not.

At this point it is convenient to introduce a new notation. Because of (8.2), the surface integrals (8.9) will not depend on the shape of the surface, but only on the singularities and their motion. Thus the surface integrals, even if they do not vanish, can be functions of τ only.

We write:

$$(8.10) \quad \frac{1}{4\pi} \int_{2l-1}^s 2\Lambda_{0r} n_r dS = \overset{s}{C}_0(\tau) = \overset{s}{C}_0$$

$$(8.11) \quad \frac{1}{4\pi} \int_{2l}^s 2\Lambda_{mr} n_r dS = \overset{s}{C}_m(\tau) = \overset{s}{C}_m$$

and assume that we have calculated the C 's. If they vanish identically, and if they vanish always as we proceed with our approximation, then our equations are self-consistent.

Let us assume, however, that the C 's in (8.10) and (8.11) are *not* zero. Then (8.1b, c) cannot be solved. There is no difficulty in solving (8.1a). This equation is of the form

$$(8.12) \quad \gamma_{00,rr} = \frac{2\Lambda_{00}}{2l-2},$$

where the right-hand side is known. We see that the solution of this equation is determined only up to an additive harmonic function. Thus we can add to any solution either single "poles" or "poles" and "dipoles."

By adding single poles we can insure the integrability of (8.1b). Then by adding dipoles we can insure the integrability of (8.1c). We could have done all that in *one* step, adding poles and dipoles, but the division into two steps makes for a simpler presentation.

After finding γ_{00} from (8.12), we calculate $\overset{s}{C}_0$ and, in general, find $\overset{s}{C}_0 \neq 0$.

We then replace in (8.1b):

$$(8.13) \quad \gamma_{00}^{2l-2} \text{ by } \gamma_{00}^{2l-2} - \sum_s 4 \dot{m}^s \psi^{2l-2}$$

where \dot{m}^s are certain functions of time to be determined soon, and ψ 's are the functions defined in (8.4). Of course this change in γ_{00}^{2l-2} induces a change in C_0^{2l-1} . Indeed,

$$(8.14) \quad \begin{cases} 2\Lambda_{0m}^{2l-1} \text{ changes now to} \\ 2\Lambda_{0m}^{2l-1} + \sum_s (4 \dot{m}^s \psi)^{2l-1}_{0m}, \end{cases}$$

as follows from (6.5b) because γ_{00}^{2l-2} appears in Λ_{0m}^{2l-1} only as $-\gamma_{00}^{2l-2} \frac{m^0}{1}$. Now obviously the old surface integral

$$(8.15) \quad \frac{1}{4\pi} \int 2\Lambda_{0r}^{2l-1} n_r dS = C_0^{2l-1}$$

changes into A.4

$$(8.16) \quad C_0^{2l-1} - 4 \dot{m}^{2l-1},$$

therefore it can be made zero by choosing

$$(8.17) \quad 4 \dot{m}^{2l-1} = C_0^{2l-1}.$$

Thus by adding a pole we can insure the integrability of (8.1b). The next step is to insure the integrability of (8.1c). Thus we assume that $\gamma_{00}^{2l-2}, \gamma_{0m}^{2l-1}$ are known, that (8.1b) is integrable and we have once more to return to γ_{00}^{2l-2} looking for a different solution of (8.1a) so as to insure the integrability of (8.1c) without destroying the integrability of (8.1b).

We replace now our γ_{00}^{2l-2} (containing the additional poles) by

$$(8.18) \quad \gamma_{00}^{2l-2} - \sum_{s=1}^p S_r^s \psi^{2l-2}_{,r}.$$

These are additional *dipole* solutions, and we assume that no other dipole expressions are contained in γ_{00}^{2l-2} . Again the S_r are functions of r only, to be determined later. The γ_{00}^{2l-2} now contain the single pole solutions so as to enforce the integrability of (8.1b). We can easily see what change in γ_{0m}^{2l-1} is induced by (8.18). The answer is, that γ_{0m}^{2l-1} changes into

$$(8.19) \quad \gamma_{0m}^{2l-1} - \sum_s (S_m^s \psi)^{2l-1}_0.$$

Indeed, if the old γ_{0m} satisfies the original equation (8.1b):

$$(8.20) \quad \gamma_{0m,ss} - \gamma_{0s,ms} = \gamma_{ms,0s} - \gamma_{00,m0} + \frac{2\Lambda'_{0m}}{2l-1},$$

then γ_{00} , γ_{0m} with the additional expressions written out in (8.18) and (8.19) satisfy the equation too. This is so, because $2\Lambda'_{0m}$ being non-linear can contain neither γ_{00} nor γ_{0m} . Therefore the addition of dipoles does not affect the integrability of (8.1b).

Now the last and decisive step: we replace in (8.1c) γ_{00} , γ_{0m} by the new expressions according to (8.18) and (8.19) and adjust the S 's so that the surface integrals will vanish identically. This requires a somewhat more lengthy calculation.

Written out explicitly, equation (8.1c) is:

$$(8.21) \quad \begin{aligned} & \gamma_{mn,ss} - \gamma_{ms,ns} - \gamma_{ns,ms} + \delta_{mn}\gamma_{rs,rs} \\ &= -\gamma_{0m,0n} - \gamma_{0n,0m} + 2\delta_{mn}\gamma_{0r,0r} + \gamma_{mn,00} - \delta_{mn}\gamma_{00,00} + \frac{2\Lambda'_{mn}}{2l} \\ &= 2\Lambda_{mn}. \end{aligned}$$

We introduce into (8.21)

$$(8.22) \quad \gamma_{00} - \sum_s \frac{S_s}{2l-2} \psi_s,$$

$$(8.23) \quad \gamma_{0m} - \sum_s \left(\frac{S_m}{2l-3} \psi_s \right)_0$$

for the old γ_{00} , γ_{0m} . We now obtain new expressions added to the old Λ_{mn} . The difficulty is, that now the contributions come not only from the linear expressions, but also from Λ'_{mn} which will contain terms of the type $\gamma_{00} \cdot \gamma_{00}$. The result of the calculations is given in A.8, and contains many expressions of which we shall here write only the first three which arise from the linear terms (the others, as we shall see, are unimportant). Instead of the old $2\Lambda_{mn}$ we have:

$$(8.24) \quad \begin{aligned} & \frac{2\Lambda_{mn}}{2l} \\ &+ \sum_s \left(\frac{S_m}{2l} \psi_s \right)_n + \frac{S_n}{2l} \psi_s \left(\frac{S_m}{2l} \psi_s \right)_m - \delta_{mn} \frac{S_r}{2l} \psi_s \left(\frac{S_r}{2l} \psi_s \right)_r \\ &+ \dots \end{aligned}$$

where the dots at the end indicate the omitted expressions. As we are here discussing the problem of surface integrals, we are justified in omitting them because they do not give any contribution to the surface integrals. We see

too, that the expressions written out here have a vanishing divergence, and this is true for the omitted terms also. Calculating the surface integrals (A.4), we find that the old surface integral

$$(8.25) \quad \frac{1}{4\pi} \int_S 2\Lambda_{mn} n_n dS = \overset{s}{C}_m$$

changes into

$$(8.26) \quad \overset{s}{C}_m - \overset{s}{S}_m.$$

Therefore it can be made zero, by choosing

$$(8.27) \quad \overset{s}{S}_m = \overset{s}{C}_m.$$

Thus we can always, by adding dipole solutions in γ_{00} , force the surface integrals to vanish identically.

By proceeding in this way, we accumulate single poles and dipoles, and the additional expressions in γ_{00} are:

$$(8.28) \quad - \sum_I \lambda^{2I-2} \sum_{s=1}^p \left(4\overset{s}{m}_{2I-2} \overset{s}{\psi} + \overset{s}{S}_{2I-2} \overset{s}{\psi}, r \right).$$

We violated our rule of not introducing dipoles. However, this was done for γ_{00} only. We can, at the end of the approximation procedure, annihilate all these additional dipole expressions by taking

$$(8.29) \quad \sum_I \lambda^{2I-2} \overset{s}{S}_r = 0.$$

Differentiating this twice, we obtain, because of (8.27):

$$(8.30) \quad \sum_I \lambda^{2I} \overset{s}{S}_m = \sum_I \lambda^{2I} \overset{s}{C}_m = 0.$$

These are the 3p equations of motion. Thus the motion is determined, if dipole solutions are rejected.

On the other hand, the m 's can be calculated from the C_0 's according to (8.17). Denoting the total coefficient at $\overset{s}{\psi}$ by $-4\overset{s}{M}$, we have:

$$(8.31) \quad \overset{s}{M} = \lambda^2 \overset{s}{m}_2 + \lambda^4 \overset{s}{m}_4 + \lambda^6 \overset{s}{m}_6 + \dots$$

where $\overset{s}{m}_2, \overset{s}{m}_4, \dots$ are functions of the original constants $\overset{s}{m}_2$ and of known functions of the time.

The equations (8.30) and (8.31) will contain only a finite number of terms depending on the order to which we wish to carry out the actual calculations.

9. On the choice of a co-ordinate system. We shall now see that it is possible to simplify our equations through the proper choice of a co-ordinate system. Let us assume that

$$(9.1) \quad \gamma_{2-2}^*, \gamma_{2-1}^*, \gamma_{2l}^*$$

are solutions of our system (6.3), where the Φ 's and Λ 's are defined by (6.4) and (6.5). Then we can show that any

$$\begin{aligned}
 (9.2) \quad & \gamma_{00} = \gamma_{00}^* \\
 & \gamma_{0m} = \gamma_{0m}^* + a_{0,m} \\
 & \gamma_{mn} = \gamma_{mn}^* + a_{m,n} + a_{n,m} - \delta_{mn} a_{r,r} + \delta_{mn} a_{0,0}
 \end{aligned}$$

with a_0, a_m arbitrary are also solutions of our equations. This can be shown just by straightforward substitution in (6.4). A simple calculation shows that all the a 's vanish from these equations. Thus we can, at each approximation step, impose four conditions upon the field. Let us choose, as is usually done, the following four co-ordinate conditions:

$$(9.3a) \quad \gamma_{00,0} - \gamma_{0r,r} = 0$$

$$(9.3b) \quad \gamma_{0m,0} - \gamma_{mr,r} = 0.$$

Indeed, if γ^* do not satisfy such a condition, then a 's can be found that ensure it. The equations for the a 's are:

$$(9.4a) \quad a_{0,rr} = \gamma_{00,0}^* - \gamma_{0r,r}^*$$

$$(9.4b) \quad a_{m,rr} = \gamma_{0m,0}^* - \gamma_{mr,r}^*.$$

With the co-ordinate condition (9.3) our system of equations is considerably simplified. Equations (6.3) now become:

$$(9.5a) \quad \gamma_{00,rr} = \gamma_{00,00} + 2\Lambda'_{00}$$

$$(9.5b) \quad \gamma_{0m,rr} = \gamma_{0m,00} + 2\Lambda'_{0m}$$

$$(9.5c) \quad \gamma_{mn,rr} = \gamma_{mn,00} + 2\Lambda'_{mn},$$

which together with the co-ordinate conditions

$$(9.6a) \quad \gamma_{00,0} - \gamma_{0r,r} = 0$$

$$(9.6b) \quad \gamma_{0m,0} - \gamma_{mr,r} = 0,$$

now form a symmetrical system of equations, where in (9.5) all the known functions on the right-hand side are at least two orders lower than those on the left.

The surface integrals that must vanish and which give the equations of motion are:

$$(9.7a) \quad \int \left(\gamma_{0m,00} - \gamma_{00,0m} + \frac{2\Lambda'_{0m}}{2l-1} \right) n_m dS = 0$$

$$(9.7b) \quad \int \left(\gamma_{nm,00} - \gamma_{n0,0m} + \frac{2\Lambda'_{nm}}{2l} \right) n_m dS = 0.$$

We can deduce them from our old formulae, using the lemma, or directly, differentiating (9.6), adding to (9.5) and using the lemma.

If, as in Sec. 8, we now introduce dipoles in order to satisfy (9.7b), we do not violate (9.6a).

Sometimes it is more convenient to use other co-ordinate conditions. For example, the one used in the actual calculations is:

$$(9.8a) \quad \gamma_{00,0} - \gamma_{00,s} = 0$$

$$(9.8b) \quad \gamma_{mn,n} = 0.$$

The equations then are:

$$(9.9a) \quad \gamma_{00,rr} = \frac{2\Lambda'_{00}}{2l-2}$$

$$(9.9b) \quad \gamma_{0m,rr} = \frac{2\Lambda'_{0m}}{2l-1}$$

$$(9.9c) \quad \gamma_{mn,rr} = -\gamma_{0m,0n} - \gamma_{0n,0m} + \delta_{mn} \gamma_{00,00} + \gamma_{mn,00} + \frac{2\Lambda'_{mn}}{2l} \\ = \frac{2\Lambda_{mn}}{2l}$$

and the surface conditions are:

$$(9.10a) \quad \int \left(\frac{2\Lambda'_{0m}}{2l-1} - \gamma_{00,0m} \right) n_m dS = 0$$

$$(9.10b) \quad \int \frac{2\Lambda_{nm}}{2l} n_m dS = 0.$$

The question arises: to what extent does the co-ordinate condition influence the equations of motion? We shall return to this problem in the last section and we shall show that the equations of motion to the sixth order do not depend on the choice of the co-ordinate system.

10. The Newtonian approximation. We shall discuss now the first three equations for $l = 2$. The equations are:

$$(10.1) \quad \gamma_{00,rr} = 0$$

$$(10.2) \quad \gamma_{0m,rr} = 0$$

$$(10.3) \quad \gamma_{nm,rr} = 2\Lambda_{nm}.$$

The co-ordinate conditions that we accept are:

$$(10.4) \quad \gamma_{0r,r} - \gamma_{00,1} = 0.$$

$$(10.5) \quad \gamma_{mr,r} = 0.$$

The explicit form of Λ_{nm} is given in A.10.

The character of our entire solution will depend essentially upon the choice of the harmonic function we take as the solution of (10.1). As we are interested in solutions representing particles, we shall write:

$$(10.6) \quad \gamma_{00} = 2\varphi; \quad \varphi = \sum_{s=1}^{\infty} \left\{ -2m\psi \right\}$$

$$\psi = \left[\left(x^r - \xi^r \right) \left(x^r - \xi^r \right) \right]^{-\frac{1}{2}} = \left(\frac{r}{r} \right)^{-1}.$$

From (10.2) we see that γ_{0m} is a harmonic function too, which must, however, satisfy the co-ordinate condition also. From (10.4) we have:

$$(10.7) \quad \gamma_{0r,r} = \gamma_{00,1} = - \sum_s \left\{ 4m \frac{\partial}{\partial x^1} \psi, \xi^r \right\}.$$

The constant $\frac{m}{2}$, which we identify with the gravitational mass of the particles is assumed to be positive. Therefore the exclusion of dipoles, together with the field equations and the co-ordinate condition determine uniquely γ_{0n} :

$$(10.8) \quad \gamma_{0n} = \sum_s 4m \psi \xi_1^n.$$

To this γ_{0n} we could add, according to (9.2) the gradient of any function and in this way obtain a general solution. But as our entire procedure consists in employing only rational functions of $\left(x^r - \xi^r \right)$, any such addition would introduce new singularities (not of the character of a single pole), or a non-Galilean field at infinity. Thus we should regard γ_{0n} in (10.8) as characterizing the problem of particles, regardless of whether we introduce the co-ordinate condition (10.4) or not.

Just for the sake of simplicity, let us now restrict our consideration to *two* particles and write (omitting the indices below \bar{m} , φ , f , g):

$$(10.9) \quad \begin{cases} \varphi = f + g \\ f = -2m\psi^1; \quad g = -2m\psi^2 \\ \xi^r = \eta^r; \quad \xi^r = \zeta^r. \end{cases}$$

The next step then, since the surface integral (9.10a) vanishes for $l = 3$, because

$$(10.10) \quad \int^2 \left(2\Lambda'_{30m} - \gamma_{00,0m}^1 \right) n_m dS = - \int^2 \gamma_{00,0m}^1 n_m dS = 0,$$

is to determine

$$(10.11) \quad \begin{aligned} C_m^1 &= \frac{1}{4\pi} \int^1 2\Lambda_{mr} n_r dS \\ C_m^2 &= \frac{1}{4\pi} \int^2 2\Lambda_{mr} n_r dS. \end{aligned}$$

If we wish to finish our approximation procedure here, the equations of motion up to the fourth, or as we shall call it, the Newtonian approximation, are:

$$(10.12) \quad C_m^1 = 0; \quad C_m^2 = 0.$$

All we have to do now is to calculate the surface integrals, according to the method outlined in A.4. The result of this particular calculation is given in A.10. It is:

$$(10.13) \quad \begin{cases} C_m^1(\tau) = 4m \left\{ \bar{\eta}^m + \frac{1}{2} \bar{g}_{,m} \right\} = 0 \\ C_m^2(\tau) = 4m \left\{ \bar{\zeta}^m + \frac{1}{2} \bar{f}_{,m} \right\} = 0 \\ \bar{g}_{,m} = g_{,m} \text{ for } x^s = \eta^s \\ \bar{f}_{,m} = f_{,m} \text{ for } x^s = \zeta^s. \end{cases}$$

The form (10.13) is actually independent of the variables x^s . In the last equations we see that $\bar{g}_{,m}$, say, is obtained by differentiating g with respect to x^s and then by replacing x^s by η^s . But the result will be the same if we *first* replace x^s by η^s and *then* differentiate with respect to η^s or ζ^s . Thus:

$$(10.14) \quad \begin{aligned} \bar{g}_{,s} &= \frac{\partial g(r)}{\partial \eta^s} = - \frac{\partial g(r)}{\partial \zeta^s} \\ g(r) &= - \frac{2m}{r}; \quad r^2 = (\eta^s - \zeta^s)(\eta^s - \zeta^s). \end{aligned}$$

We can, therefore, think of our equations of motion as involving the differentiation of functions depending only on the position of singularities, as is characteristic of the theories based on the concept of action at a distance. Indeed, we see that our equations are precisely the Newtonian equations of motion, deduced here as the first approximation from the field equations. The treatment of p particles (instead of two) does not add any new difficulties if we deal with the Newtonian approximation only.

11. Transition to the next approximation. We wish to go now beyond the Newtonian approximation. But then we must calculate γ_{mn} , since Λ_{mn} depends on γ_{mn} . The characteristic feature of this method is that generally, if we wish to find the equations of motion to the $2l$ approximation (inclusive) then we do not need to calculate γ_{mn} , because C_m does not contain it. But now, if we wish to go one step further we must find γ_{mn} for which the equations are:

$$(11.1) \quad \gamma_{mn,rr} = 2\Lambda_{mn}.$$

This is "the transition step" that we have to take before proceeding to the next approximation. These equations are integrable only if we *do* assume Newtonian motion. Otherwise we would have to add dipoles. Yet if we wish to proceed *only* to the next approximation we may assume Newtonian motion and additional expressions induced by the dipole fields are not necessary.

If in (11.1) we assume Newtonian motion, then (11.1) can be integrated, because the surface integral of Λ_{mn} vanishes then. But if we do this, we introduce Newtonian motion into Λ_{mn} . This is admissible because any difference between Λ calculated this way and Λ calculated with the proper motion is of order Λ . Thus since we do not propose to go beyond Λ we may ignore the additional dipole fields. It is for this reason that the previous special calculations in [1] were correct, but the general theory was not.

We shall now solve

$$(11.2) \quad \gamma_{mn,rr} = 2\Lambda_{mn} = -\gamma_{0m,0n} - \gamma_{0n,0m} + 2\delta_{mn}\varphi_{,00} \\ - 2\varphi\varphi_{,mn} - \varphi_{,m}\varphi_{,n} + \frac{2}{3}\delta_{mn}\varphi_{,s}\varphi_{,s}$$

assuming the Newtonian equation of motion, i.e. (10.13).

We can ignore the dipole expressions because we are interested only in the equations of motion to the next approximation. But, for the same reason, we are interested only in those expressions in γ_{mn} which give a contribution to the corresponding surface integral of Λ_{mn} .

An inspection of Λ_{mn} (A.12) shows that we need only the knowledge of γ_{mn} in the neighbourhood of the singularities, and we may ignore in it the terms which do not become infinite as $r \rightarrow 0$, since the surface integral due to these terms must vanish (see A.12). On the other hand γ_{ss} which also appears in Λ should and will be calculated in the entire space.

In the equation (11.2) we have, on the right-hand side "cross products," that is, products belonging to different singularities. Because of them (11.2) can only be integrated in the neighbourhood of the first singularity, say. The expression arising from the second singularity can be expanded into a power series near the first singularity. Retaining all the expressions that may give some contribution to the surface integral and those only, we have in the neighbourhood of the first singularity:

$$(11.3) \quad \left\{ \begin{aligned} \gamma_{mn} = & \left\{ f[(x^n - \eta^n)\eta^m + (x^m - \eta^m)\eta^n - \delta_{mn}(x^s - \eta^s)\eta^s] \right\}, 0 \\ & + \left\{ g[(x^n - \xi^n)\xi^m + (x^m - \xi^m)\xi^n - \delta_{mn}(x^s - \xi^s)\xi^s] \right\}, 0 \\ & + \frac{1}{4} r^2 f_{,m} f_{,n} + \frac{1}{4} r^2 g_{,m} g_{,n} \\ & - f_{,m}(x^n - \eta^n) \bar{g} \\ & + \alpha_{mn} f + \beta_{mn} g. \end{aligned} \right.$$

Here only the expression

$$- f_{,m}(x^n - \eta^n) \bar{g}; \quad \bar{g} = g \text{ for } x^s = \eta^s,$$

is due to the interaction terms. The two last expressions are the additive harmonic functions (dipoles are excluded) and they are determined by the co-ordinate condition

$$(11.4) \quad \gamma_{mr,r} = 0.$$

The result is:

$$(11.5) \quad \left\{ \begin{aligned} \alpha_{mn} &= 2\eta^m \eta^n + \delta_{mn} \bar{g} \\ \beta_{mn} &= 2\xi^m \xi^n + \delta_{mn} \bar{f}, \\ \bar{f} &= f(r); \quad \bar{g} = g(r); \quad r^2 = (\eta^s - \xi^s)(\eta^s - \xi^s). \end{aligned} \right.$$

But, let us say once more, that all this is true only if the Newtonian motion is assumed.

Finally, as we mentioned before, γ_{rr} can be calculated rigorously. The result is:

$$(11.6) \quad \gamma_{rr} = - 2m \frac{1}{r},_{00} - 2m^2 \frac{1}{r},_{00} + \frac{1}{4} \varphi^2 + \alpha f + \beta g.$$

Here the α and β are determined so that near the singularity (11.6) will be consistent with (11.3) for $m = n = r$. The result is:

$$(11.7) \quad \left\{ \begin{aligned} \alpha &= 2\eta^s \eta^s + \frac{1}{2} \bar{g} \\ \beta &= 2\xi^s \xi^s + \frac{1}{2} \bar{f}. \end{aligned} \right.$$

Thus our transition steps are accomplished.

12. Beyond the Newtonian approximation. We write down the next field equations:

$$(12.1a) \quad \gamma_{00,rr} = \frac{2\Lambda_{00}}{4} = -\frac{3}{2} \varphi_{,r} \varphi_{,r}$$

$$(12.1b) \quad \gamma_{0m,rr} = \frac{2\Lambda'_{0m}}{5} = \varphi_{,s} \gamma_{0s,m} - \varphi_{,sm} \gamma_{0s} - 3\varphi_{,r} \varphi_{,rm}$$

$$(12.1c) \quad \gamma_{mn,rr} = \frac{2\Lambda_{mn}}{6}.$$

The explicit expressions for Λ_{mn} are quoted in A.12. The solution of (12.1a) is simple:

$$(12.2) \quad \gamma_{00} = -\frac{3}{4} \varphi^2 - \frac{4}{4} m \dot{\varphi} - \frac{4}{4} m \ddot{\varphi}.$$

As we know from the general theory, the arbitrary harmonic functions have to be determined in such a way as to make (12.1b) self-consistent, that is, the corresponding surface integral must vanish.

The co-ordinate conditions, are here, as before,

$$(12.3a) \quad \gamma_{0r,r} - \gamma_{00,0} = 0$$

$$(12.3b) \quad \gamma_{mr,r} = 0.$$

Because of this, the conditions for solvability of (12.1b, c) are:

$$(12.4a) \quad \frac{1}{4\pi} \int \left\{ \frac{2\Lambda'_{0m}}{5} - \gamma_{00,0m} \right\} n_m dS = 0$$

$$(12.4b) \quad \frac{1}{4\pi} \int \frac{2\Lambda_{mr}}{6} n_r dS = 0.$$

We have in (12.4a) the equations that determine m . The result of evaluating the surface integrals in (12.4a), (see A.12) is:

$$(12.5) \quad \begin{cases} \frac{1}{4} m = \frac{1}{2} \frac{1}{m} \left\{ \dot{\eta}^s \dot{\eta}^s + \frac{1}{2} \ddot{\xi} \right\} = \frac{1}{2} \left(\frac{1}{m} \dot{\eta}^s \dot{\eta}^s - \frac{1}{m} \frac{2}{m} \frac{1}{r} \right) \\ \frac{2}{4} m = \frac{1}{2} \frac{2}{m} \left\{ \dot{\xi}^s \dot{\xi}^s + \frac{1}{2} \ddot{\eta} \right\} = \frac{1}{2} \left(\frac{2}{m} \dot{\xi}^s \dot{\xi}^s - \frac{1}{m} \frac{2}{m} \frac{1}{r} \right) \\ \frac{1}{m} = \frac{1}{m} ; \frac{2}{m} = \frac{2}{m} ; r^2 = (\eta^s - \xi^s)(\eta^s - \xi^s). \end{cases}$$

The next step, after the self-consistency of (12.1b) has been insured is to calculate the γ_{0s} . We need them, because they enter into the next surface

integral. Including only relevant terms that can influence the surface integral we find near the first singularity:

$$(12.6) \quad \left\{ \begin{aligned} \gamma_{0m} = & -\frac{1}{4} r f_{,m} f_{,r} \dot{\eta}^r + \frac{3}{4} f^2 \dot{\eta}^m \\ & + \frac{3}{2} (x^s - \eta^s) (\dot{\eta}^s - \dot{\zeta}^s) f \dot{g}_{,m} \\ & - (x^m - \eta^m) f \dot{g}_{,s} (\dot{\eta}^s - \dot{\zeta}^s) \\ & + \frac{1}{2} (x^s - \eta^s) f_{,m} \dot{\zeta}^s \{ \dot{g} + \dot{g}_{,r} (x^r - \eta^r) \} \\ & + \frac{1}{2} (x^s - \eta^s) \{ f \dot{g}_{,s} \dot{\zeta}^m + f_{,m} \dot{g} \dot{\zeta}^s \} \\ & + a_{0m} f. \end{aligned} \right.$$

Again a_{0m} is determined from the co-ordinate condition (12.3a) and the result is:

$$(12.7) \quad a_{0m} = -\dot{\eta}^s \dot{\eta}^s \dot{\eta}^m + \dot{g} \dot{\eta}^m - \dot{g} \dot{\zeta}^m.$$

Now the scene is set for the last and most difficult calculation:

$$(12.8) \quad C_m = \frac{1}{4\pi} \int_0^1 2\lambda_{mn} n_m dS.$$

Some remarks about this calculation are made in A.12, and partial results given. We obtain:

$$\begin{aligned} \frac{1}{6} C_m = & -\frac{1}{4\pi m} \left\{ \left[\dot{\eta}^s \dot{\eta}^s + \frac{3}{2} \dot{\zeta}^s \dot{\zeta}^s - 4\dot{\eta}^s \dot{\zeta}^s - 4\frac{\dot{m}}{r} - 5\frac{\dot{m}}{r} \right] \frac{\partial}{\partial \eta^m} \left(\frac{1}{r} \right) \right. \\ & \left. + [4\dot{\eta}^s (\dot{\zeta}^m - \dot{\eta}^m) + 3\dot{\eta}^m \dot{\zeta}^s - 4\dot{\zeta}^s \dot{\zeta}^m] \frac{\partial}{\partial \eta^s} \left(\frac{1}{r} \right) + \frac{1}{2} \frac{\partial^2 r}{\partial \eta^s \partial \eta^r \partial \eta^m} \dot{\zeta}^s \dot{\zeta}^r \right\}. \end{aligned}$$

Thus the equation of motion belonging to this stage of approximation is:

$$(12.10) \quad \lambda^s C_m + \lambda^s C_m = 0.$$

We can now re-absorb the λ 's by substituting new units for τ and $\frac{1}{m}, \frac{2}{m}$:

old $\tau = \lambda \cdot$ new τ ; old mass $= \lambda^{-1} \cdot$ new mass.

Preserving the old symbols for the new units we have for the equations of motion of the first particle:

$$\begin{aligned} \ddot{\eta}^m - \frac{2}{m} \frac{\partial(1/r)}{\partial \eta^m} = & \frac{2}{m} \left\{ \left[\dot{\eta}^s \dot{\eta}^s + \frac{3}{2} \dot{\zeta}^s \dot{\zeta}^s - 4\dot{\eta}^s \dot{\zeta}^s - 4\frac{\dot{m}}{r} - 5\frac{\dot{m}}{r} \right] \frac{\partial}{\partial \eta^m} (1/r) \right. \\ (12.11) \quad & + [4\dot{\eta}^s (\dot{\zeta}^m - \dot{\eta}^m) + 3\dot{\eta}^m \dot{\zeta}^s - 4\dot{\zeta}^s \dot{\zeta}^m] \frac{\partial}{\partial \eta^s} (1/r) \\ & \left. + \frac{1}{2} \frac{\partial^2 r}{\partial \eta^s \partial \eta^r \partial \eta^m} \dot{\zeta}^s \dot{\zeta}^r \right\}. \end{aligned}$$

The equations of motion for the other particle are obtained by replacing

$$\overset{1}{m}, \overset{2}{m}, \eta, \zeta, \text{ by } \overset{2}{m}, \overset{1}{m}, \zeta, \eta,$$

respectively.

These are the equations of motion of two particles. They can be integrated and conclusions concerning perihelion motion of a double star can be drawn from them [5]. The entire method can also be adapted for the case of a charged particle in an electromagnetic field [4].

13. The equations of motion and the co-ordinate condition. The contents of the last three sections are not new. Its presentation, however, is different than that given before in [1] and [2], since it has been adjusted to the new theory. There is one more question that we wish to answer and which we did not treat before. It is possible to do so only now after the general theory has been perfected. We ask: To what extent do the equations of motion as formulated in (12.11) depend on the particular choice of the co-ordinate system?

We reject any particular choice of co-ordinate system and write the first two equations:

$$(13.1) \quad \Phi_{00} + 2\Lambda_{00} = -\gamma_{00,rr} = 0$$

$$(13.2) \quad \Phi_{0m} + 2\Lambda_{0m} = -\gamma_{0m,rr} + \gamma_{0r,mr} - \gamma_{00,m} \overset{1}{1} = 0.$$

We assume that we start our approximation procedure with the same γ_{00} and γ_{0m} functions as we did before. But from now on, while dealing with the rest of the equations we shall look for *general* solutions not restricted by any additional co-ordinate conditions.

Thus the equations that we wish to consider now are:

$$(13.3a) \quad \Phi_{mn} + 2\Lambda_{mn} = 0$$

$$(13.3b) \quad \Phi_{00} + 2\Lambda_{00} = 0$$

$$(13.3c) \quad \Phi_{0m} + 2\Lambda_{0m} = 0.$$

In the previous three sections we solved these equations, using special co-ordinate conditions. Let us now call the special solutions that we obtained there:

$$(13.4) \quad \gamma_{mn}^*, \gamma_{00}^*, \gamma_{0m}^*.$$

Knowing them, as we do, we can find the general solution of (13.3). The procedure is similar to that outlined in Sec. 9, only slightly different, because we have now a set of equations of order $(2l)$, $(2l)$, and $(2l + 1)$, whereas before

we had a set of order $(2l - 2)$, $(2l - 1)$, and $(2l)$. But a straightforward substitution shows, that because of the linear expressions in (13.3), (and they alone enter the argument), the general solution of (13.3) is:

$$(13.5a) \quad \gamma_{mn} = \gamma_{mn}^* + a_{m,n} + a_{n,m} - \delta_{mn} a_{r,r}$$

$$(13.5b) \quad \gamma_{00} = \gamma_{00}^* + a_{r,r}$$

$$(13.5c) \quad \gamma_{0m} = \gamma_{0m}^* + a_{0,m} + a_{m,0}$$

where a_μ are arbitrary. The question then is: If we substitute these new expressions into the Λ 's do we change the integrals

$$(13.6) \quad \int_4 \Lambda_{mr} n_r dS, \int_5 \Lambda_{0r} n_r dS, \int_6 \Lambda_{mr} n_r dS?$$

As far as the first two integrals are concerned the answer is easy; Λ_4 is not changed; only linear expressions in Λ_5 are affected, but the surface integral of the additional expressions disappears because of the lemma. But it is different with the third surface integral. In Λ_6 new terms appear containing the a 's. They appear both through the linear and the non-linear expressions. But these additional expressions—quoted in the last appendix—are such that their surface integral vanishes. Thus in the sense explained here the equations of motion do not depend on the choice of the co-ordinate system. This dependence would appear probably in the next approximation steps (Λ_8), but it does not enter into the surface integral of Λ_6 . This is a satisfying result, because it is difficult to see the meaning of our co-ordinate conditions

$$(13.7) \quad \begin{aligned} \gamma_{mr,r} &= 0 \\ \gamma_{0r,r} - \gamma_{00,0} &= 0 \\ \gamma_{mr,r} &= 0 \end{aligned}$$

and it is good to know that our equations of motion are independent of it. This result is general. If we have a system

$$(13.8) \quad \begin{aligned} \Phi_{mn} + 2\Lambda_{mn} &= 0 \\ \Phi_{00} + 2\Lambda_{00} &= 0 \\ \Phi_{0m} + 2\Lambda_{0m} &= 0, \end{aligned}$$

then the surface integral of Λ_{mn} is independent of the co-ordinate conditions introduced in this particular approximation stage. This is so, because the a 's combine with the φ 's in the same way in each approximation step.

APPENDICES

A.2

The field equations are:

$$(A.2, 1) \quad R_{\mu\nu} = - \left\{ \frac{\rho}{\mu\nu} \right\}_{|\rho} + \left\{ \frac{\rho}{\mu\rho} \right\}_{|\nu} + \left\{ \frac{\rho}{\mu\sigma} \right\} \left\{ \frac{\sigma}{\rho\nu} \right\} - \left\{ \frac{\rho}{\mu\nu} \right\} \left\{ \frac{\sigma}{\rho\sigma} \right\}.$$

Introducing here the h 's as defined in (2.3) and splitting (A.2, 1) into linear and non-linear terms we have

$$(A.2, 2a) \quad R_{00} = -\frac{1}{2} h_{00|ss} + h_{0s|0s} - \frac{1}{2} h_{ss|00} + L'_{00}$$

$$(A.2, 2b) \quad R_{0n} = -\frac{1}{2} h_{0n|ss} + \frac{1}{2} h_{0s|ns} + \frac{1}{2} h_{ns|0s} - \frac{1}{2} h_{ss|n0} + L'_{0n}$$

$$(A.2, 2c) \quad R_{mn} = -\frac{1}{2} h_{mn|ss} + \frac{1}{2} h_{ms|ns} + \frac{1}{2} h_{ns|ms} - \frac{1}{2} h_{ss|mn} + \frac{1}{2} h_{mn|00} - \frac{1}{2} h_{m0|n0} - \frac{1}{2} h_{n0|m0} + \frac{1}{2} h_{00|mn} + L'_{mn}.$$

Here $L'_{\mu\nu}$ are the non-linear expressions. We form now:

$$(A.2, 3) \quad -2(R_{\mu\nu} - \frac{1}{2} \eta_{\mu\nu} \eta^{\alpha\beta} R_{\alpha\beta}) = 0.$$

Substituting the γ 's for the h 's, we see that (A.2, 3) written out is (2.10)—(2.18), where

$$(A.2, 4) \quad \Lambda'_{\mu\nu} = L'_{\mu\nu} - \frac{1}{2} \eta_{\mu\nu} \eta^{\alpha\beta} L_{\alpha\beta}.$$

A.4

In calculating the surface integrals we need to take into account only expressions that go to infinity like r^{-2} , because only such expressions will give finite contributions. Since all the field functions are finite (outside of the singularity), and since the contributions do not depend on the shape of the surface, we may ignore all other expressions. But we have to keep the surface fixed, because in our calculations a complicated expression whose surface integral does not depend on the shape of the surface, is split into partial expressions with non-vanishing divergence. Thus in our calculations the surface is always a two-dimensional "sphere" with radius shrinking to zero. Let us assume, for the sake of simplicity, that the space co-ordinate of the singularity is (0, 0, 0). We shall first give some examples of the surface integrals formed around such a singularity.

Example 1. We calculate:

$$\int_0^\infty \psi_{,s} n_s dS; \quad \psi = r^{-1}; \quad r^2 = x^s x^s,$$

We have:

$$\int_0^\infty \psi_{,s} n_s dS = - \int \frac{x^s x^s}{r^4} r^2 \sin \theta d\theta d\varphi = -4\pi.$$

Example 2. We calculate:

$$\int_0^\infty \psi_{,s} n_r dS = - \int \frac{x^s x^r}{r^4} r^2 \sin \theta d\theta d\varphi = -\frac{4\pi}{3} \delta_{sr}.$$

Example 3. We calculate:

$$\int_0^0 \psi_{,mn} n_n \chi(r) dS.$$

To find such a surface integral we expand $\chi(r)$ as a power series in the neighbourhood of the singularity:

$$\chi = \chi(0) + \chi_{,s}(0)x_s + \dots$$

The only contribution is from the second expression, that is, we have to calculate:

$$\begin{aligned} \chi_{,s}(0) \int_0^0 \psi_{,mn} n_n x^s dS &= -\chi_{,s}(0) \int_0^0 \frac{x^m x^s}{r^4} r^3 \sin \theta d\theta d\varphi \\ &\quad + 3\chi_{,s}(0) \int_0^0 \frac{x^m x^s}{r^4} r^3 \sin \theta d\theta d\varphi \\ &= \frac{8\pi}{3} \chi_{,m}(0). \end{aligned}$$

In the course of our calculations we shall have to find more complicated surface integrals and the following table will prove to be useful:

Table of Surface Integrals

- I. $\frac{1}{4\pi} \int_0^0 \psi_{,n} n_n dS = -1.$
- II. $\frac{1}{4\pi} \int_0^0 \psi_{,s} n_n dS = -\frac{1}{3} \delta_{sn}.$
- III. $\frac{1}{4\pi} \int_0^0 x^r \psi_{,n} n_n = \frac{2}{3} \delta_{rs}.$
- IV. $\frac{1}{4\pi} \int_0^0 x^r \psi_{,ms} n_n dS = -\frac{1}{15} \{2\delta_{rn} \delta_{ms} - 3\delta_{rm} \delta_{ns} - 3\delta_{rs} \delta_{mn}\}.$
- V. $\frac{1}{4\pi} \int_0^0 x^r \psi_{,rs} n_n dS = \frac{2}{3} \delta_{ns}.$
- VI. $\frac{1}{4\pi} \int_0^0 x^n \psi_{,ms} n_n dS = 0.$
- VII. $\frac{1}{4\pi} \int_0^0 x^n x^s \psi_{,mr} n_n = 0.$
- VIII. $\frac{1}{4\pi} \int_0^0 x^m x^s \psi_{,nr} n_n dS = \frac{2}{3} \delta_{ms} \delta_{lr} - \frac{2}{3} (\delta_{ml} \delta_{rs} + \delta_{mr} \delta_{ls}).$

A.8

The linear terms of (8.26) give the following contribution to $2\Lambda_{mn}$:

$$(A.8, 1) \quad \sum_{s=1}^2 \left\{ S_m^s \psi_{,n} + S_n^s \psi_{,m} - \delta_{mn} S_r^s \psi_{,r} \right\}, \quad \frac{00}{2}.$$

The non-linear terms can be found in the following way: Inspecting the terms in Λ_{mn} (A.12, 3) we see products of γ_{00} and γ_{0n} or, as it is there called 2φ . Thus, if we put there the expression in (8.18) in place of γ_{00} and write for brevity:

$$(A.8, 2) \quad (S_r\psi) = \sum_i \sum_{2l-2}^i S_i \psi$$

we get five new terms. Thus with the abbreviation (A.8, 2) we have in every approximation the following additional terms:

$$(A.8, 3) \quad \left\{ \begin{aligned} & \{ (S_m\psi)_{,n} + (S_n\psi)_{,m} - \delta_{mn}(S_r\psi)_{,r} \}_{,00} \\ & + \varphi(S_r\psi)_{,rmn} + \frac{1}{2} \varphi_{,n}(S_r\psi)_{,rm} \\ & + \frac{1}{2} \varphi_{,m}(S_r\psi)_{,rn} - \frac{3}{2} \delta_{mn} \varphi_{,s}(S_r\psi)_{,rs} + \varphi_{,mn}(S_r\psi)_{,r} \end{aligned} \right.$$

Only three of the linear terms give us a contribution to the surface integral. It is more difficult to see that the non-linear terms do not give any contribution, since it requires some knowledge of how to deal with surface integrals which is outlined in A.4, and which we shall here assume. We can write the non-linear terms in (A.8, 3), in the following way:

$$(A.8, 4) \quad \begin{aligned} & \{ \varphi_{,mn}(S_r\psi) \}_{,r} - \{ \varphi_{,mr}(S_n\psi) \}_{,r} \\ & + \varphi_{,mn}(S_n\psi)_{,r} \\ & + \frac{1}{2} \{ \varphi_{,n}(S_r\psi)_{,m} \}_{,r} - \frac{1}{2} \{ \varphi_{,r}(S_n\psi)_{,m} \}_{,r} \\ & + \frac{1}{2} \varphi_{,r}(S_n\psi)_{,mr} \\ & - \frac{3}{2} \delta_{mn} \varphi_{,r}(S_s\psi)_{,sr} \end{aligned}$$

These are the non-linear expressions, and their divergence vanishes because φ is a harmonic function. The expressions written out in pairs in (A.8, 4) do not give any contribution to the surface integrals, because of our lemma in Sec. 3. Thus the only contribution could come from the terms:

$$(A.8, 5) \quad \frac{3}{2} \varphi_{,s} S_n \psi_{,ms} - \frac{3}{2} \delta_{mn} \varphi_{,s} S_r \psi_{,rs}.$$

Here only the "cross products" could give contributions and we find with the help of the table in A.4, that the result is zero.

A.10

In the $l = 2$ approximation we have:

$$(A.10, 1) \quad \left\{ \begin{aligned} \gamma_{00} &= 2\varphi = 2f + 2g \\ \gamma_{0n} &= -2f\eta^n - 2g\xi^n = h_{0n} \\ h_{00} &= \varphi = f + g \\ h_{00}^{(0)} &= -h_{00} = -\varphi \\ h_{0n}^{(0)} &= h_{0n} = \gamma_{0n} \\ h_{mn} &= -h^{mn} = \delta_{mn}\varphi. \end{aligned} \right.$$

A straightforward calculation gives:

$$(A.10, 2) \quad \left\{ \begin{array}{l} 2\Lambda_{00} = 0 \\ 2\Lambda_{0m} = -\frac{\gamma_{00,m0}}{2} \\ 2\Lambda_{mn} = -\frac{\gamma_{0m,0n}}{3} - \frac{\gamma_{0n,0m}}{3} + 2\delta_{mn}\varphi_{,00} \\ \quad - 2\varphi\varphi_{,mn} - \varphi_{,m}\varphi_{,n} + \frac{3}{2}\delta_{mn}\varphi_{,s}\varphi_{,s} \end{array} \right.$$

The contributions to the surface integrals are (for the first singularity):

$$\begin{aligned} -\gamma_{0m,0n} &\rightarrow 4\frac{1}{m}\ddot{\eta}^m.4\pi \\ -\gamma_{0n,0m} &\rightarrow 4\frac{1}{m}\ddot{\eta}^m.4\pi \\ 2\delta_{mn}\varphi_{,00} &\rightarrow -4\frac{1}{m}\ddot{\eta}^m.4\pi \\ -2\varphi\varphi_{,mn} &\rightarrow -\frac{8}{3}\frac{1}{m}\ddot{g}_{,m}.4\pi \\ -\varphi_{,m}\varphi_{,n} &\rightarrow -\frac{8}{3}\frac{1}{m}\ddot{g}_{,m}.4\pi \\ \frac{3}{2}\delta_{mn}\varphi_{,s}\varphi_{,s} &\rightarrow 2\frac{1}{m}\ddot{g}_{,m}.4\pi \\ &(\dot{m} = \frac{1}{2}). \end{aligned}$$

A.12

A straightforward calculation of Λ , Λ , Λ gives

$$(A.12, 1) \quad 2\Lambda_{00} = -\frac{3}{2}\varphi_{,s}\varphi_{,s}$$

$$(A.12, 2) \quad 2\Lambda'_{0m} = \varphi_{,s}\gamma_{0s,m} - \varphi_{,sm}\gamma_{0s} - 3\varphi_{,s}\varphi_{,m}$$

$$\begin{aligned} 2\Lambda_{mn} = & -\frac{\gamma_{0m,0n}}{5} - \frac{\gamma_{0n,0m}}{5} + \frac{\delta_{mn}\gamma_{00,00}}{4} + \frac{\gamma_{mn,00}}{4} - \frac{\varphi\gamma_{00,mn}}{4} \\ & - \frac{\varphi\gamma_{ss,mn}}{4} - \frac{\varphi_{,mn}\gamma_{00}}{4} - \frac{\varphi_{,mn}\gamma_{ss}}{4} + \frac{\varphi_{,ms}\gamma_{ns}}{4} \\ & + \frac{\varphi_{,ns}\gamma_{ms}}{4} - \frac{\delta_{mn}\varphi_{,sr}\gamma_{sr}}{4} - 2\frac{\varphi_{,s}\gamma_{mn,s}}{4} + \frac{\varphi_{,s}\gamma_{ms,s}}{4} \\ & + \frac{\varphi_{,s}\gamma_{ns,s}}{4} - \frac{1}{2}\frac{\varphi_{,m}\gamma_{ss,n}}{4} - \frac{1}{2}\frac{\varphi_{,n}\gamma_{ss,m}}{4} - \frac{1}{2}\frac{\varphi_{,n}\gamma_{00,m}}{4} \\ & - \frac{1}{2}\frac{\varphi_{,m}\gamma_{00,n}}{4} + \frac{3}{2}\frac{\delta_{mn}\varphi_{,s}\gamma_{rr,s}}{4} + \frac{3}{2}\frac{\delta_{mn}\varphi_{,s}\gamma_{00,s}}{4} \\ & - \frac{\gamma_{0s}\gamma_{0n,ms}}{3} - \frac{\gamma_{0s}\gamma_{0m,ns}}{3} + 2\frac{\gamma_{0s}\gamma_{0s,mn}}{3} \\ (A.12, 3) \quad & + \frac{1}{2}\frac{\delta_{mn}\gamma_{0s,r}\gamma_{0r,s}}{3} - \frac{3}{2}\frac{\delta_{mn}\gamma_{0s,r}\gamma_{0s,r}}{3} + \frac{\gamma_{0s,m}\gamma_{0s,n}}{3} \\ & + \frac{\gamma_{0m,s}\gamma_{0n,s}}{3} - \frac{\varphi_{,0n}\gamma_{0m}}{3} - \frac{\varphi_{,0m}\gamma_{0n}}{3} + 2\frac{\delta_{mn}\gamma_{0s}\varphi_{,0s}}{3} \\ & - \frac{\varphi_{,0}\gamma_{0m,n}}{3} - \frac{\varphi_{,0}\gamma_{0n,m}}{3} - \frac{\varphi_{,n}\gamma_{0m,0}}{3} - \frac{\varphi_{,m}\gamma_{0n,0}}{3} \\ & + 2\frac{\varphi\gamma_{0m,0n}}{3} + 2\frac{\varphi\gamma_{0n,0m}}{3} - 2\frac{\delta_{mn}\varphi\varphi_{,00}}{3} \\ & + 2\varphi\varphi_{,mn} - \varphi\varphi_{,m}\varphi_{,n} + \frac{3}{2}\delta_{mn}\varphi\varphi_{,s}\varphi_{,s} \\ & + \frac{1}{2}\delta_{mn}\varphi_{,s}\varphi_{,s} \end{aligned}$$

The surface integral (12.4a) for $s = 1$ is, because of (12.2), and (12.1b):

$$\frac{1}{4\pi} \int (\varphi, r \gamma_{0r, m} - \varphi, r m \gamma_{0n} - \frac{3}{2} \varphi, s \varphi, m + \frac{3}{2} \varphi \varphi, s m + 4 \frac{1}{4} (m \psi)_{,0m}) n_m dS = 0.$$

The contributions of these five expressions are respectively:

$$(1) \rightarrow -\frac{4m}{3} \tilde{g}_{,s} \dot{s}^s - 4m \tilde{g}_{,s} \dot{\eta}^s$$

$$(2) \rightarrow -\frac{8m}{3} \tilde{g}_{,s} \dot{s}^s$$

$$(3) \rightarrow 3m \tilde{g}_{,s} \dot{s}^s + m \tilde{g}_{,s} \dot{\eta}^s$$

$$(4) \rightarrow 2m \tilde{g}_{,s} \dot{\eta}^s$$

$$(5) \rightarrow -4m \frac{1}{4}.$$

Therefore:

$$\begin{aligned} -4m \frac{1}{4} &= m \tilde{g}_{,s} \dot{s}^s + m \tilde{g}_{,s} \dot{\eta}^s = 2m \tilde{g}_{,s} \dot{\eta}^s - m \tilde{g}_{,s} \dot{\eta}^s + m \tilde{g}_{,s} \dot{s}^s \\ &= -m(2\dot{\eta}^s \dot{\eta}^s + \tilde{g})_{,0}. \end{aligned}$$

From the last equation (12.5) follows immediately.

The last step is to calculate the surface integrals due to Λ . Here a skilful use of the lemma may save the calculation of many surface integrals. Indeed, $2\Lambda_{mn}$ can be written in the following form:

$$\begin{aligned} 2\Lambda_{mn} &= (\varphi, n \gamma_{sm} - \varphi, s \gamma_{nm})_{,s} + (\varphi \gamma_{ms, n} - \varphi \gamma_{mn, s})_{,s} \\ &+ (\delta_{ms} \varphi, r \gamma_{rn} - \delta_{mn} \varphi, r \gamma_{rs})_{,s} + (\delta_{mn} \varphi, s \gamma_{rr} - \delta_{ms} \varphi, n \gamma_{rr})_{,s} \\ &+ \frac{1}{2} (\delta_{mn} \varphi \gamma_{rr, s} - \delta_{ms} \varphi \gamma_{rr, n})_{,s} + (\delta_{mn} \gamma_{0s, 0} - \delta_{ms} \gamma_{0n, 0})_{,s} \\ &+ \frac{1}{2} (\gamma_{0s, m} \gamma_{0n} - \gamma_{0n, m} \gamma_{0s})_{,s} + (\delta_{mn} \varphi, s \gamma_{0s} - \delta_{ms} \varphi, s \gamma_{0n})_{,s} \\ &+ (\gamma_{0n} \gamma_{0m, s} - \gamma_{0s} \gamma_{0m, n})_{,s} + \frac{1}{2} (\delta_{mn} \gamma_{0s, r} \gamma_{0r} - \delta_{ms} \gamma_{0n, r} \gamma_{0r})_{,s} \\ &+ (\delta_{ms} \gamma_{0r, n} \gamma_{0r} - \delta_{mn} \gamma_{0r, s} \gamma_{0r})_{,s} \\ &- \gamma_{0m, 0n} + \gamma_{mn, 00} + \gamma_{0s} \gamma_{0s, mn} & [a_1 + a_2 + a_3] \\ &- \frac{1}{2} \delta_{mn} \gamma_{0s, r} \gamma_{0s, r} - (\varphi, n \gamma_{0m})_{,0} & [a_4 + a_5] \\ &- (\varphi, m \gamma_{0n})_{,0} + (\varphi \gamma_{0m, n})_{,0} + (\varphi \gamma_{0n, m})_{,0} & [a_6 + a_7 + a_8] \\ &- \frac{3}{2} \delta_{mn} \varphi, s \varphi, s - \frac{1}{2} \varphi \gamma_{ss, mn} & [a_9 + a_{10}] \\ &+ \frac{1}{2} \varphi, n \gamma_{ss, m} + \frac{1}{2} \varphi, n \gamma_{00, m} & [a_{11} + a_{12}] \\ &+ \frac{1}{2} \varphi, m \gamma_{00, n} - \frac{1}{2} \delta_{mn} \varphi, s \gamma_{00, s} & [a_{13} + a_{14}] \\ &- \varphi \varphi, 00 \delta_{mn} - 2\varphi \varphi, m \varphi, n & [a_{15} + a_{16}] \\ &+ \frac{1}{4} \varphi \varphi, s \varphi, s \delta_{mn}. & [a_{17}] \end{aligned}$$

TABLE OF SURFACE INTEGRALS FOR $\int_0^1 \Delta_{ms} \eta_s dS$

No.	Expression	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	a_{16}	a_{17}	Result	Remarks
1	$\frac{1}{m} \tilde{g}_{,s} \eta^s \eta^m$	$\frac{16}{-3}$			$\frac{4}{-3}$		$\frac{4}{-3}$	$\frac{8}{-3}$	$\frac{4}{-15}$	$\frac{4}{-15}$	$\frac{8}{15}$	$\frac{4}{15}$				$\frac{4}{5}$			-8	$\tilde{g}_{,s} = -2\tilde{m} \frac{\partial^1}{\partial \eta^s}$
2	$\frac{1}{m} \tilde{g}_{,m} \eta^m$	-2					$\frac{4}{-3}$	-4	$\frac{4}{-3}$			$\frac{29}{-3}$	$\frac{3}{3}$	$\frac{11}{3}$	$\frac{5}{-3}$	$\frac{2}{3}$	$\frac{32}{3}$	$\frac{22}{-3}$	-8	$\tilde{g}_{,m} = -\frac{2\tilde{m}}{r}; \eta^m = -\frac{1}{3} \tilde{g}_{,m}$
3	$\frac{1}{m} \tilde{g}_{,m} \eta^s \eta^s$	1					$\frac{4}{-3}$	$\frac{4}{-5}$			$\frac{8}{5}$	$\frac{4}{5}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{-3}$	$\frac{4}{-15}$			2	$\tilde{g}_{,m} = -2\tilde{m} \eta^m$
4	$\frac{1}{m} \tilde{g}_{,ms} \eta^s \eta^s$											2	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{-3}$				3	
5	$\frac{1}{m} \tilde{g}_{,m} \tilde{f}$	$\frac{4}{3}$			2	$\frac{2}{3}$						$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{-6}$				5	$\tilde{g}_{,m} \tilde{f} = -\tilde{g}_{,m}; \tilde{f} = -\frac{2\tilde{m}}{r}$
6	$\frac{1}{m} \tilde{g}_{,m} \tilde{r}_{,0em}$											-2							-2	$\tilde{r}_{,0em} = (\tilde{r}_{,0em})$ for $x^s = \eta^s$
7	$\frac{1}{m} \tilde{g}_{,s} \eta^s \eta^s$	$\frac{16}{5}$			$\frac{8}{3}$	$\frac{4}{5}$		$\frac{4}{3}$											8	
8	$\frac{1}{m} \tilde{g}_{,s} \eta^s \eta^m$	$\frac{16}{5}$				$\frac{8}{-15}$		4	$\frac{4}{3}$	-2									6	
9	$\frac{1}{m} \tilde{g}_{,m} \eta^s \eta^s$	$\frac{32}{-15}$		$\frac{16}{-3}$		$\frac{4}{-5}$		4	$\frac{4}{-3}$										-8	
10	$\frac{1}{m} \tilde{g}_{,s} \eta^s \eta^m$	$\frac{8}{-3}$			-4	$\frac{4}{-3}$													-8	

$$\bullet \tilde{r}_{,0em} = \frac{\partial^2 \tilde{r}}{\partial \eta^s \partial \eta^m} \tilde{r}^{s,m}, \text{ as } \frac{\partial^2 \tilde{r}}{\partial \eta^s \partial \eta^m} = 0.$$

Because of the lemma we have to find now the surface integrals of only 17 expressions denoted successively by a_1, a_2, \dots, a_{17} . The result of this calculation is summarized in the table. Only ten types of expressions (or their equivalents) appear in the result. The table tells us what is the contribution of each of the a 's to the final result. The only a that does not give a contribution is $a_2 = \gamma_{mn,00}$.

A.13

The additional expressions in Λ_{mn} induced through rejection of the coordinate condition are:

$$\begin{aligned} & 2 (\delta_{mn} a_{0,r0} - \delta_{mr} a_{0,n0}),_r \\ & + (\varphi_{,m} a_{n,r} - \varphi_{,n} a_{r,m}),_r \\ & + (\varphi_{,n} a_{m,r} - \varphi_{,r} a_{m,n}),_r \\ & + (\varphi_{,n} a_{s,m} - \varphi_{,s} a_{n,m}),_s \\ & - 2(\delta_{mn} \varphi_{,s} a_{s,r} - \delta_{mr} \varphi_{,s} a_{s,n}),_r \\ & + 2(\delta_{mn} \varphi_{,s} a_{r,r} - \delta_{ms} \varphi_{,n} a_{r,r}),_r. \end{aligned}$$

They are written in such a way, that the vanishing of each line is evident, because of the lemma.

REFERENCES

- [1] A. Einstein, L. Infeld and B. Hoffmann, *Ann. of Math.*, vol. 39, 1 (1938) 66.
- [2] A. Einstein and L. Infeld, *Ann. of Math.*, vol. 41, 2 (1940) 455.
- [3] L. Infeld, *Phys. Rev.* vol. 53 (1938) 836.
- [4] L. Infeld and P. R. Wallace, *Phys. Rev.*, vol. 57, (1940) 797.
- [5] H. P. Robertson, *Ann. of Math.*, vol. 39, 1 (1938) 101.

Institute for Advanced Study
University of Toronto

SOME PROPERTIES OF THE EIGENFUNCTIONS OF THE LAPLACE-OPERATOR ON RIEMANNIAN MANIFOLDS

S. MINAKSHISUNDARAM AND Å. PLEIJEL

Introduction. Let V be a connected, compact, differentiable Riemannian manifold. If V is not closed we denote its boundary by S . In terms of local coordinates (x^i) , $i = 1, 2, \dots, N$, the line-element dr is given by¹

$$dr^2 = g_{ik}(x^1, x^2, \dots, x^N) dx^i dx^k$$

where $g_{ik}(x^1, x^2, \dots, x^N)$ are the components of the metric tensor on V . We denote by Δ the Beltrami-Laplace-Operator

$$\Delta u = \frac{1}{\sqrt{g}} \frac{\partial}{\partial x^i} \left(\sqrt{g} g^{ik} \frac{\partial u}{\partial x^k} \right)$$

and we consider on V the differential equation

$$(1) \quad \Delta u + \lambda u = 0.$$

If V is closed this equation will in general have an infinite number of eigenvalues $\lambda = \lambda_m$, $m = 1, 2, \dots$, and corresponding eigenfunctions $\phi_m(P)$ where P is a point in V . When V has a boundary we have to consider in addition to (1), certain boundary-conditions in order to define eigenvalue-problems. We consider the following conditions, either

$$(2) \quad u = 0 \text{ on } S,$$

or

$$(3) \quad \frac{\partial u}{\partial n} = 0 \text{ on } S,$$

where $\frac{\partial}{\partial n}$ denotes a differentiation in the direction of the normal of S . Eigenvalues and eigenfunctions shall be denoted in the same way as was indicated in the case of a closed manifold. We assume in all cases the eigenvalues to have been arranged in non-decreasing order of magnitude and the eigenfunctions to form a complete orthonormal set

$$\int_V \phi_i(P) \phi_b(P) dV = \delta_{ib}$$

($dV = \sqrt{g} dx^1 dx^2 \dots dx^N$). In the problem with a closed manifold and in (1), (3) the value $\lambda_0 = 0$ is a simple eigenvalue ($\lambda_1 > 0$) with the corresponding eigenfunction equal to the constant $\frac{1}{\sqrt{V}}$ where V denotes the volume of the manifold. In the problem (1), (2) we have $\lambda_0 > 0$.

Received July 4, 1948.

¹We use the usual notations of tensor-calculus (g is the determinant of the covariant metric tensor g_{ik}).

We always assume V together with its boundary (if this exists) to be sufficiently regular so that those theorems from the theory of eigenvalue-problems which are required are valid.

The aim of this note is to study the analytic continuations in the s -plane of the Dirichlet's series (summation from $m = 0$ to $+\infty$ or from $m = 1$ to $+\infty$ according as $\lambda_0 > 0$ or $= 0$)

$$(4) \quad \sum \frac{\phi_m(P)\phi_m(Q)}{\lambda_m^s},$$

and

$$(5) \quad \sum \frac{\phi_m^2(P)}{\lambda_m^s},$$

where if V has a boundary, P and Q shall be interior points of V . In the case where V is a bounded two-dimensional Euclidean domain, the series (5) was first studied by Carleman [1]. Later in the case where V is a bounded Euclidean domain of arbitrary dimension N it was shown by Minakshisundaram [7] by a method different from Carleman's that (4) is an entire function of s with zeros at negative integers and that (5) is a meromorphic function with simple pole at $s = \frac{1}{2}N$ and zeros at negative integers. The method here developed is a generalization of Carleman's.

Even though our results are valid to a certain extent under less restrictive regularity conditions it is convenient to state them here for an analytic manifold V . If $\lambda_0 > 0$ (the formulation of the results is only slightly different in the case when $\lambda_0 = 0$) we find that both the series (4), (5) can be continued arbitrarily far to the left of their abscissas of convergence. The continuation of (4) is an entire function with zeros at non-positive integers while (5) represents a function analytic except for simple poles at

$$s = \frac{1}{2}N - \nu, \nu = 0, 1, 2, \dots \quad \text{if } N \text{ is odd,}$$

and at

$$s = \frac{1}{2}N, \frac{1}{2}N - 1, \frac{1}{2}N - 2, \dots, 2, 1 \quad \text{if } N \text{ is even.}$$

The residue at the poles can be determined in terms of the g_{ik} . If N is odd the function defined by (5) has zeros at non-positive integers and if N is even its values in these points can be explicitly determined from the metric tensor of V . By Ikehara's theorem (see [14], p. 44) we obtain as a corollary the relation

$$\sum_{\lambda_m < T} \phi_m^2(P) \sim \frac{T^{N/2}}{(2\sqrt{\pi})^N \Gamma\left(\frac{N}{2} + 1\right)},$$

where the sign \sim indicates that the quotient of both the sides tends to 1 when T tends to $+\infty$; see [1].

In the case of a closed manifold the series

$$(6) \quad \sum_{m=1}^{\infty} \lambda_m^{-s}$$

is easily seen to have properties similar to those stated for (5) and by the help

of Ikehara's theorem we obtain immediately the asymptotic distribution of the eigenvalues. In the case when V has a boundary our method does not give such complete results concerning (6). It is possible by generalizing Carleman's method to deduce the asymptotic eigenvalue-distribution also in this case, but it seems as if this could be done more easily by already available methods (see [3] and [13]). The analytic continuation of (4), (5) and (6) in the case of a sphere was previously studied by Minakshisundaram [8].

1. Construction of a parametrix. We introduce normal coordinates in the neighbourhood of an inner point P of V . If r_{PQ} denotes the geodesic distance from P to Q and $\frac{d}{dr}$ differentiation along a geodesic from P , the normal coordinates (y^i) of Q are defined by

$$y^i = r_{PQ} \left(\frac{dx^i}{dr} \right)_{Q=P}.$$

If Φ is a function of $r = r_{PQ}$ only and U is an arbitrary function we observe that

$$(7) \quad \Delta \Phi U = U \left(\frac{d^2 \Phi}{dr^2} + \frac{N-1}{r} \frac{d\Phi}{dr} + \frac{d \log \sqrt{g}}{dr} \frac{d\Phi}{dr} \right) + 2 \frac{dU}{dr} \frac{d\Phi}{dr} + \Phi \Delta U$$

on using the well-known formulae

$$r_{PQ}^2 = g_{ik}(P) y^i y^k, \\ g_{ik}(Q) y^k = g_{ik}(P) y^k,$$

where the fundamental tensor is determined with respect to the normal coordinates and (y^i) are the coordinates of the point Q .

We define in a neighbourhood of P

$$H_n(P, Q; t) = \frac{1}{(2\sqrt{\pi})^n} e^{-\frac{r^2}{4t}} t^{-\frac{N}{2}} (U_0 + U_1 t + \dots U_n t^n)$$

by choosing $U_\nu(P, Q)$, $\nu = 0, 1, 2, \dots, n$, independent of t and solutions of the differential equations ($U_{-1} \equiv 0$)

$$r \frac{dU_\nu}{dr} + \frac{r}{2} \frac{d \log \sqrt{g}}{dr} U_\nu + \nu U_\nu = \Delta U_{\nu-1}.$$

The functions $U_\nu(P, Q)$ are uniquely determined by the conditions that they shall be finite for $P = Q$ and by the normalizing condition $U_0(P, P) = 1$. It is apparent that the choice of the integer n is limited by the regularity of V . However, we shall assume V so regular that n can be chosen $> \frac{1}{2}N - 2$ (see Sec. 2). We find

$$U_0(P, Q) = \left(\frac{g(P)}{g(Q)} \right)^{\frac{1}{4}}$$

and for $\nu > 0$

$$U_\nu(P, Q) = \frac{U_0(P, Q)}{r_{PQ}} \int_P^Q \frac{r_{P\Pi}^{\nu-1} \Delta_\Pi U_{\nu-1}(P, \Pi)}{U_0(P, \Pi)} dr_{P\Pi}.$$

By the help of (7) we find, on account of the definition of U_ν , that

$$(8) \quad \left(\Delta - \frac{\partial}{\partial t} \right) H_n = \frac{\Delta U_n}{(2\sqrt{\pi})^n} e^{-\frac{r^2}{4t}} t^{-\frac{N}{2}+n}.$$

For $Q = P$, $t = 0$ the singularity of $H_n(P, Q; t)$ coincides with the singularity of a fundamental solution of the heat-equation

$$\Delta u - \frac{\partial u}{\partial t} = 0.$$

$H_n(P, Q; t)$ is a *parametrix* of this equation.

By use of a Laplace-transformation we obtain from $H_n(P, Q; t)$ the function

$$K_n(P, Q; -\xi) = \frac{1}{(2\sqrt{\pi})^n} \sum_{\nu=0}^n U_\nu \int_0^\infty e^{-\xi t - \frac{r^2}{4t}} t^{-\frac{N}{2}+\nu} dt.$$

The singularity of this function for $Q = P$ coincides with the singularity of a fundamental solution of the equation

$$(9) \quad \Delta u - \xi u = 0.$$

From (8) it follows that

$$(10) \quad (\Delta - \xi) K_n(P, Q; -\xi) = \frac{\Delta U_n}{(2\sqrt{\pi})^n} \int_0^\infty e^{-\xi t - \frac{r^2}{4t}} t^{-\frac{N}{2}+n} dt.$$

The function $K_n(P, Q; -\xi)$ is a *parametrix* of (9).

The construction of H_n (and K_n) is analogous to Hadamard's construction of a fundamental solution of the General Wave Equation (see [6]). But by comparing our construction with Hadamard's we see that Hadamard's proof of the convergence of the infinite series he considers cannot be used for the infinite series we obtain from $H_n(P, Q; t)$ by letting n tend to infinity. The same is seen to be true *a fortiori* for the infinite series obtained from

$$K_n(P, Q; -\xi).$$

As a *parametrix in the large* we consider

$$\Gamma_n(P, Q; -\xi) = \eta_R(r_{PQ}) K_n(P, Q; -\xi)$$

where $\eta_R(r)$ is a continuous function of r satisfying

$$\eta_R(r) = \begin{cases} 1 & \text{when } r \leq \frac{R}{2}, \\ 0 & \text{when } r \geq R, \end{cases}$$

and having continuous derivatives of order one and two. In the interval $\frac{1}{2}R \leq r \leq R$ the function $\eta_R(r)$ can be chosen as a polynomial satisfying inequalities of the form

$$\left| \eta_R \right| \leq 1, \quad \left| \frac{d\eta_R}{dr} \right| \leq \frac{\text{const.}}{R}, \quad \left| \frac{d^2\eta_R}{dr^2} \right| \leq \frac{\text{const.}}{R^2}.$$

R shall be chosen so small that the geodesic sphere round P with radius R is contained in the neighbourhood of P where the construction of $K_n(P, Q; -\xi)$ is valid. In the case of a closed manifold V we can choose R independent of P .

In the case of a manifold with a boundary the choice of R must depend on the distance from P to this boundary.

2. The Green's function. With the aid of the parametrix $\Gamma_n(P, Q; -\xi)$ we may express the Green's function of (9) in the form

$$(11) \quad G(P, Q; -\xi) = \Gamma_n(P, Q; -\xi) - \gamma_n(P, Q; -\xi).$$

The integer n shall be chosen so large, $n > \frac{1}{2}N - 2$, that all singularities of $G(P, Q; -\xi)$ are contained in $\Gamma_n(P, Q; -\xi)$ and $\gamma_n(P, P; -\xi)$ is finite. As a function of the point Q the "regular part" of the Green's function, viz., $\gamma_n(P, Q; -\xi)$ satisfies the equation

$$(\Delta - \xi)_Q \gamma_n(P, Q; -\xi) = (\Delta - \xi)_Q \Gamma_n(P, Q; -\xi).$$

If V has a boundary, $G(P, Q; -\xi)$ shall satisfy the prescribed boundary condition (2) or (3). On account of the vanishing of Γ_n and $\frac{\partial \Gamma_n}{\partial n}$ on S the function $\gamma_n(P, Q; -\xi)$ satisfies the same boundary condition as the Green's function itself.

Now $\gamma_n(P, Q; -\xi)$ can be obtained as a solution of the variational problem: to search for the minimum of (ξ is supposed to be real and positive)

$$E(u; \Gamma_n) = \int_V \left(g^{ik} \frac{\partial u}{\partial x^i} \frac{\partial u}{\partial x^k} + \xi u^2 + 2Fu \right) dV$$

where $F(Q) = (\Delta - \xi)_Q \Gamma_n(P, Q; -\xi)$. The admissible functions u are assumed to be continuous with piece-wise, continuous, first-order derivatives in the open kernel of V . The integral

$$\int_V \left(g^{ik} \frac{\partial u}{\partial x^i} \frac{\partial u}{\partial x^k} + \xi u^2 \right) dV$$

shall be finite. If the boundary condition $u = 0$ on S is considered, the admissible functions shall also satisfy this condition (see [4], p. 482).

The minimizing function $\gamma_n(P, Q; -\xi)$ satisfies the equation (see [10])

$$(12) \quad \gamma_n(P, P; -\xi) = E(\gamma_n; \Gamma_n) - \int_V \Gamma_n(\Delta \Gamma_n - \xi \Gamma_n) dV$$

from which we observe that on the one hand

$$\gamma_n(P, P; -\xi) \leq - \int_V \Gamma_n(\Delta \Gamma_n - \xi \Gamma_n) dV,$$

and on the other

$$\begin{aligned} \gamma_n(P, P; -\xi) &\geq - \frac{1}{\xi} \int_V F^2 dV - \int_V \Gamma_n(\Delta \Gamma_n - \xi \Gamma_n) dV \\ &= - \frac{1}{\xi} \int_V (\Delta \Gamma_n - \xi \Gamma_n)^2 dV - \int_V \Gamma_n(\Delta \Gamma_n - \xi \Gamma_n) dV. \end{aligned}$$

The first inequality follows from the fact that $u \equiv 0$ is an admissible function

in our minimum-problem, the second is obtained by forming a complete square under the integral-sign in $E(u; \Gamma_n)$. Thus an estimate for $\gamma_n(P, P; -\xi)$ for large positive values of ξ can be deduced from the estimate for

$$(13) \quad \int_V \Gamma_n(\Delta \Gamma_n - \xi \Gamma_n) dV$$

and

$$(14) \quad \frac{1}{\xi} \int_V (\Delta \Gamma_n - \xi \Gamma_n)^2 dV.$$

By help of the estimations for (13), (14) obtained in the next paragraph we find

$$(15) \quad |\gamma_n(P, P; -\xi)| \leq \text{const. } \xi^{\frac{N}{2}-n-2}$$

where the constant depends on R .

3. Auxiliary estimations. On account of the inequalities

$$|\eta_R| \leq 1, \quad \frac{1}{2} \left(\xi t + \frac{r^2}{4t} \right) \geq \frac{r\sqrt{\xi}}{2}$$

and

$$(16) \quad C^{-\frac{N}{2}} \sum_{s=0}^n |U_s| t^s \leq \text{const.}, \quad (\xi \geq \xi_0 > 0),$$

it follows when $N > 2$ that

$$(17) \quad |\Gamma_n(P, Q; -\xi)| \leq \text{const. } C^{-\frac{r\sqrt{\xi}}{2}} \int_0^\infty C^{-\frac{r^2}{4t}} t^{\frac{N}{2}} dt = \text{const. } \frac{C^{-\frac{r\sqrt{\xi}}{2}}}{r^{\frac{N-2}{2}}}.$$

When $N = 2$ we use instead of (16) the inequality

$$C^{-\frac{N}{2}} \sum_{s=0}^n |U_s| t^s \leq \text{const. } C^{-\alpha t}, \quad (\xi \geq \xi_0 > 0; 0 < \alpha < \frac{1}{2}),$$

and by the help of well-known properties of the Bessel-function (see [12], pp. 183, 80, 202)

$$K_0(z) = \frac{1}{2} \int_0^\infty e^{-z \frac{\tau^2}{4\tau}} \tau^{-1} d\tau$$

for z tending to 0 and to $+\infty$ we obtain in this case

$$(18) \quad |\Gamma_n(P, Q; -\xi)| \leq \text{const. } e^{-\frac{r\sqrt{\xi}}{2}} \log(r\sqrt{\xi}).$$

For (10) we see that for $r \leq \frac{1}{2}R$

$$(19) \quad |\Delta \Gamma_n - \xi \Gamma_n| \leq \text{const. } e^{-\frac{r\sqrt{\xi}}{2}} \int_0^\infty e^{-\frac{t}{2}} t^{-\frac{N}{2}+n} dt = \text{const. } \xi^{\frac{N}{2}-n-1} e^{-\frac{r\sqrt{\xi}}{2}}$$

provided $-\frac{1}{2}N + n > -1$. When $-\frac{1}{2}N + n \leq -1$ we use the method which gave us (17) and (18) and deduce in this way:

$$(20) \quad |\Delta\Gamma_n - \xi\Gamma_n| \leq \text{const. } r^{-N+2n+2} e^{-\frac{r\sqrt{\xi}}{2}} \text{ when } -\frac{1}{2}N + n < -1,$$

and

$$(21) \quad |\Delta\Gamma_n - \xi\Gamma_n| \leq \text{const. } e^{-\frac{r\sqrt{\xi}}{2}} \log(r\sqrt{\xi}) \text{ when } -\frac{1}{2}N + n = -1.$$

In the interval $\frac{1}{2}R \leq r \leq R$ we find inequalities for $|\Delta\Gamma_n - \xi\Gamma_n|$ in which the majorizing expressions contain the factor $e^{-\frac{R\sqrt{\xi}}{2}}$. This fact and the inequality $r \geq \frac{1}{2}R$ make it possible to give to these inequalities the same forms as (19), (20), (21) the constants being now dependent on R .

By introducing (17) and (19) in the expressions (13) and (14) we find

$$\left| \int_V \Gamma_n (\Delta\Gamma_n - \xi\Gamma_n) dV \right| \leq \text{const. } \xi^{\frac{N}{2}-n-2},$$

and

$$(22) \quad \left| \frac{1}{\xi} \int_V (\Delta\Gamma_n - \xi\Gamma_n)^2 dV \right| \leq \text{const. } \xi^{\frac{N}{2}-2n-3}$$

on account of the fact that dV can approximately be substituted by the Euclidean volume-element $r^{N-1} dr d\Omega$. Observing (see the beginning of Sec. 2) that $-N + 2n + 4$ is a positive integer it is easily seen that all combinations of the inequalities (17), (18) with (19), (20), (21) give the same result (22).

4. A fundamental formula. Starting from the relation

$$G(P, Q; -\xi) - G(P, Q; -\xi_0) = -(\xi - \xi_0) \int_V G(P, \Pi; -\xi) G(\Pi, G; -\xi_0) dV_\Pi$$

we obtain by repeated application

$$(23) \quad G(P, Q; -\xi) = \sum_{p=0}^p (-1)^p (\xi - \xi_0)^p G^{(p)}(P, Q; -\xi_0) \\ = (-1)^{p+1} (\xi - \xi_0)^{p+1} \int_V G(P, \Pi; -\xi) G^{(p)}(\Pi, Q; -\xi_0) dV_\Pi,$$

where

$$G^{(0)}(P, Q; -\xi_0) = G(P, Q; -\xi_0), \\ G^{(k+1)}(P, Q; -\xi_0) = \int_V G(P, \Pi; -\xi_0) G^{(k)}(\Pi, Q; -\xi_0) dV_\Pi.$$

We assume without detailed discussion that the integral

$$\int_V (G^{(q)}(P, Q; -\xi_0))^2 dV_Q$$

is finite when² $q \geq \left\lceil \frac{N}{4} \right\rceil$ and P is an interior point of V . In the case when V is a bounded Euclidean domain this follows from the inequality

$$|G(P, Q; -\xi_0)| \leq \frac{\text{const.}}{r_{PQ}^{\frac{N-2}{2}}}.$$

² $[a] = \text{integer, } a - 1 < [a] \leq a.$

For other cases we may refer to the works of Giraud [5] and de Rham [11] (for the case when V is closed). It follows that for $p \geq 2 \left[\frac{N}{4} \right]$ the right side of

$$(24) \quad \Gamma_n(P, Q; -\xi) - \gamma_n(P, Q; -\xi) - \sum_{r=0}^p (-1)^r (\xi - \xi_0)^r G^{(r)}(P, Q; -\xi_0) \\ = (-1)^{p+1} (\xi - \xi_0)^{p+1} \sum_{m=0}^{\infty} \frac{\phi_m(P) \phi_m(Q)}{(\lambda_m + \xi)(\lambda_m + \xi_0)^{p+1}}.$$

Since the right side of this relation is finite for $Q = P$ the singular parts of $\Gamma_n(P, Q; -\xi)$ must cancel the singular parts of the sum

$$(25) \quad \sum_{r=0}^p (-1)^r (\xi - \xi_0)^r G^{(r)}(P, Q; -\xi_0).$$

What remains after a transition to the limit, $Q \rightarrow P$, is the "finite part" of $\Gamma_n(P, Q; -\xi)$ for $Q = P$ minus $\gamma_n(P, P; -\xi)$ minus a polynomial in ξ of degree p contributed by the finite part for $Q = P$ of the sum (25). In order to calculate the finite part of $\Gamma_n(P, Q; -\xi)$ for $Q = P$ we have to consider the finite contributions for $r = 0$ from

$$\int_0^{\infty} e^{-4t-\frac{r^2}{4t}} t^{-\frac{N}{2}+r} dt = 2^{\frac{N}{2}-r} \left(\frac{\sqrt{\xi}}{r} \right)^{\frac{N}{2}-r-1} K_{\frac{N}{2}-r-1}(r\sqrt{\xi}),$$

where $K_{\zeta}(z)$ denotes the Bessel K -function of order ζ (see [12], p. 183). If N is odd, $2\zeta = N - 2r - 2$ is odd and ([12], p. 78)

$$K_{\zeta}(z) = \frac{\pi}{2 \sin \pi \zeta} (I_{-\zeta}(z) - I_{\zeta}(z)).$$

By this formula and by well-known developments of the Bessel functions $I_{-\zeta}(z)$ and $I_{\zeta}(z)$ we find that the finite contribution from $\Gamma_n(P, Q; -\xi)$ for $Q = P$ is, when N is odd

$$(26) \quad M_n(\xi, P) = \frac{1}{(2\sqrt{\pi})^N} \sum_{r=0}^n \Gamma\left(-\frac{N}{2} + r + 1\right) \xi^{\frac{N}{2}-r-1} U_r(P, P).$$

When N is even we make use of similar considerations but now based on the formula (ζ is a positive integer; see [12], pp. 79 and 80)

$$K_{\pm \zeta}(z) = \frac{1}{2} \left(\frac{z}{2} \right)^{\zeta} \sum_{m=0}^{\zeta-1} (-1)^m \frac{\Gamma(\zeta - m)}{m!} \left(\frac{z}{2} \right)^{2m} + (-1)^{\zeta+1} \left(\frac{z}{2} \right)^{\zeta} \\ \cdot \sum_{m=\zeta}^{\infty} \frac{\left(\frac{z}{2} \right)^{2m}}{m! \Gamma(\zeta + m + 1)} \left(\log \frac{z}{2} - \frac{1}{2} \psi(m+1) - \frac{1}{2} \psi(\zeta + m + 1) \right)$$

where

$$\psi(k+1) = 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{k} - \gamma,$$

$$\psi(1) = -\gamma,$$

γ = Euler's constant.

The finite contribution from $\Gamma_n(P, Q; -\xi)$ for $Q = P$ obtained in this case is

$$(27) \quad M_n(\xi, P) = \frac{1}{(2\sqrt{\pi})^N} \sum_{\nu=0}^{\frac{N}{2}-1} (-1)^{\frac{N}{2}-\nu} \frac{\xi^{\frac{N}{2}-\nu-1}}{\Gamma\left(\frac{N}{2}-\nu\right)} \left(\log \xi - 2 \log 2 \right. \\ \left. - \psi(1) - \psi\left(\frac{N}{2}-\nu\right) \right) + \frac{1}{(2\sqrt{\pi})^N} \sum_{\nu=\frac{N}{2}}^n \Gamma\left(-\frac{N}{2}+\nu+1\right) \xi^{\frac{N}{2}-\nu-1}.$$

So we have from (24)

$$(28) \quad M_n(\xi, P) - \gamma_n(P, P; -\xi) + A_p(\xi, \xi_0, P) \\ = (-1)^{p+1} (\xi - \xi_0)^{p+1} \sum_{m=0}^{\infty} \frac{\phi_m^2(P)}{(\lambda_m + \xi)(\lambda_m + \xi_0)^{p+1}},$$

where $A_p(\xi, \xi_0, P)$ is a polynomial of degree p in ξ with coefficients depending on ξ_0 and P and where $M_n(\xi, P)$ is equal to (26) or (27) according as N is odd or even. In the case when $\lambda_0 > 0$ we obtain by performing the transition to the limit $\xi_0 \rightarrow 0$ that

$$(29) \quad M_n(\xi, P) - \gamma_n(P, P; -\xi) + \sum_{\nu=0}^p A_\nu(P) \xi^\nu \\ = (-1)^{p+1} (\xi - \xi_0)^{p+1} \sum_{m=0}^{\infty} \frac{\phi_m^2(P)}{(\lambda_m + \xi) \lambda_m^{p+1}}.$$

When $\lambda_0 = 0$ we transfer the first term of the series on the right in (28) to the left, take it together with $A_n(\xi, \xi_0, P)$ and note that since the other terms of the equation remain finite for $\xi_0 \rightarrow 0$, the value of

$$A_p(\xi, \xi_0, P) - (-1)^{p+1} (\xi - \xi_0)^{p+1} \frac{\phi_0^2(P)}{\xi \xi_0^{p+1}}$$

remains finite and gives an expression of the form

$$\sum_{\nu=0}^p A_\nu(P) \xi^\nu - \frac{1}{V\xi}$$

V being the volume of the manifold. We obtain in the case when $\lambda_0 = 0$ the formula

$$(30) \quad M_n(\xi, P) - \gamma_n(P, P; -\xi) + \sum_{r=0}^p A_r(P) \xi^r - \frac{1}{V\xi} \\ = (-1)^{p+1} \xi^{p+1} \sum_{m=1}^{\infty} \frac{\phi_m^2(P)}{(\lambda_m + \xi) \lambda_m^{p+1}}.$$

5. Analytic continuation of (5). We suppose first $\lambda_0 > 0$ and multiply both sides of (29) by

$$\frac{1}{2\pi i(-\xi)^s} = \frac{1}{2\pi i} e^{-s \log |\xi| - is(\arg \xi - \pi)}$$

and integrate along the following contour in the complex ξ -plane with a cut along the real positive axis. From $+\infty$ to a (a real, $0 < a < \lambda_0$) along the lower part of the cut, from a to a along a circle round the origin and then from a to $+\infty$ along the upper part of the cut. We obtain in this way when the real part $\Re s$ of s is sufficiently large

$$(31) \quad \sum_{m=0}^{\infty} \frac{\phi_m^2(P)}{\lambda_m^s} = \frac{\sin \pi s}{\pi} \left(- \int_a^{\infty} \frac{M_n(\xi, P)}{\xi^s} d\xi + \int_a^{\infty} \frac{\gamma_n(P, P; -\xi)}{\xi^s} d\xi \right. \\ \left. + \sum_{r=0}^p \frac{A_r(P) a^{r-s+1}}{s-r-1} \right) - \frac{a^{1-s} e^{i\pi s}}{2\pi} \int_0^{2\pi} F(ae^{i\theta}, P) e^{i\theta(1-s)} d\theta$$

where

$$F(\xi, P) = (-1)^{p+1} \xi^{p+1} \sum_{m=0}^{\infty} \frac{\phi_m^2(P)}{(\lambda_m + \xi) \lambda_m^{p+1}}.$$

The last integral in (31) is an entire function of s vanishing when s equals a non-positive integer and so is the expression

$$\frac{\sin \pi s}{\pi} \sum_{r=0}^p \frac{A_r(P) a^{r-s+1}}{s-r-1}.$$

On account of our estimation for $\gamma_n(P, P; -\xi)$, (15), the integral

$$\int_a^{\infty} \frac{\gamma_n(P, P; -\xi)}{\xi^s} d\xi$$

is a regular function in the half-plane $\Re s > \frac{1}{2}N - n - 2$. In the case when N is odd the first integral in (31) is equal to the following expression (see (26))

$$(32) \quad - \int_a^{\infty} \frac{M_n(\xi, P)}{\xi^s} d\xi = \frac{1}{(2\sqrt{\pi})^N} \sum_{r=0}^n \Gamma\left(-\frac{N}{2} + r + 1\right) \frac{a^{\frac{N}{2}-r-s}}{s - \frac{N}{2} + r} U_r(P, P)$$

and in the case when N is even (see (27))

$$\begin{aligned}
 & - \int_a^{\infty} \frac{M_n(\xi, P)}{\xi^s} d\xi = \frac{1}{(2\sqrt{\pi})^N} \sum_{\nu=0}^{\frac{N}{2}-1} \frac{(-1)^{\frac{N}{2}-\nu} a^{\frac{N}{2}-\nu-s}}{\Gamma\left(\frac{N}{2}-\nu\right)\left(s-\frac{N}{2}+\nu\right)} \\
 (33) \quad & \cdot \left(\log a + \frac{1}{s-\frac{N}{2}+\nu} - 2 \log 2 - \psi(1) - \psi\left(\frac{N}{2}-\nu\right) \right) U_\nu(P, P) \\
 & + \frac{1}{(2\sqrt{\pi})^N} \sum_{\nu=\frac{N}{2}}^{\infty} \Gamma\left(-\frac{N}{2}+\nu+1\right) \frac{a^{\frac{N}{2}-\nu-s}}{s-\frac{N}{2}+\nu} U_\nu(P, P).
 \end{aligned}$$

All taken together we obtain from (31), (32), (33) the following

THEOREM. *The Dirichlet's series*

$$(34) \quad \sum_{m=0}^{\infty} \frac{\phi_m^2(P)}{\lambda_m^s}$$

can be continued to the left of its abscissa of absolute convergence. The function $\zeta(s, P)$ thus obtained can be written

$$\zeta(s, P) = \frac{1}{(2\sqrt{\pi})^N} \sum_{\nu=0}^{\infty} \frac{U_\nu(P, P)}{\Gamma\left(\frac{N}{2}-\nu\right)\left(s-\frac{N}{2}+\nu\right)} + R_n(s, P) \text{ if } N \text{ is odd,}$$

and

$$\zeta(s, P) = \frac{1}{(2\sqrt{\pi})^N} \sum_{\nu=0}^{\frac{N}{2}-1} \frac{U_\nu(P, P)}{\Gamma\left(\frac{N}{2}-\nu\right)\left(s-\frac{N}{2}+\nu\right)} + R_n(s, P) \text{ if } N \text{ is even,}$$

where in both cases $R_n(s, P)$ is regular in the half-plane $\Re s > \frac{1}{2}N - n - 2$. When s is equal to a non-positive integer ($> \frac{1}{2}N - n - 2$) we have

$$\zeta(s, P) = 0 \text{ in the case when } N \text{ is odd,}$$

and

$$\zeta(s, P) = \frac{\Gamma(1-s)}{(2\sqrt{\pi})^N} U_{\frac{N}{2}-s}(P, P) \text{ when } N \text{ is even.}$$

In the case when $\lambda_0 = 0$ we have to consider the series

$$(35) \quad \sum_{m=1}^{\infty} \frac{\phi_m^2(P)}{\lambda_m^s}$$

instead of (34) and to use the relation (30) instead of (29). We obtain essentially

the same theorem but because of the term $-\frac{1}{V\xi}$ in (30) we arrive at the relations

$$\zeta(0, P) = -\frac{1}{V} \text{ when } N \text{ is odd,}$$

and

$$\zeta(0, P) = \frac{U_N(P, P)}{(2\sqrt{\pi})^N} - \frac{1}{V} \text{ when } N \text{ is even,}$$

instead of the corresponding relation in the theorem (we denote here by $\zeta(s, P)$ the analytic continuation of (35)).

In so far as the Dirichlet's series (34) or (35) with positive terms can be continued to the left of its abscissa of absolute convergence with a simple pole at $s = \frac{1}{2}N$ and residue (we observe that $U_0(P, P) = 1$)

$$\frac{1}{(2\sqrt{\pi})^N \Gamma\left(\frac{N}{2}\right)}$$

we could apply Ikehara's theorem (see [18], p. 44) and obtain the following asymptotic distribution of the squares of the eigenfunctions as a

$$\text{COROLLARY.} \quad \sum_{\lambda_m \leq T} \phi_m^2(P) \sim \frac{T^{\frac{N}{2}}}{(2\sqrt{\pi})^N \Gamma\left(\frac{N}{2} + 1\right)}.$$

6. Analytic continuation of (4). Let P and Q be two different (inner) points of V and let R be so chosen that the geodesic spheres round P and Q with radius R have no points in common. From the formula (24) we obtain when $\lambda_0 > 0$

$$\begin{aligned} (36) \quad \Gamma_n(P, Q; -\xi) - \gamma_n(P, Q; -\xi) + \sum_{r=0}^p B_r(P, Q) \xi^r \\ = (-1)^{p+1} \xi^{p+1} \sum_{m=0}^{\infty} \frac{\phi_m(P) \phi_m(Q)}{(\lambda_m + \xi) \lambda_m^{p+1}}. \end{aligned}$$

By choice of R the value of $\Gamma_n(P, Q; -\xi)$ is zero and we have only to estimate $\gamma_n(P, Q; -\xi)$ as a function of ξ for $\xi \rightarrow +\infty$. Corresponding to (12) we have

$$\begin{aligned} (37) \quad & 2\{\gamma_n(P, Q; -\xi) + \gamma_n(Q, P; -\xi)\} \\ & = E(\gamma_n(P, \Pi; -\xi) + \gamma_n(Q, \Pi; -\xi); \Gamma_n(P, \Pi; -\xi) + \Gamma_n(Q, \Pi; -\xi)) \\ & - E(\gamma_n(P, \Pi; -\xi) - \gamma_n(Q, \Pi; -\xi); \Gamma_n(P, \Pi; -\xi) - \Gamma_n(Q, \Pi; -\xi)) \\ & + 2\{\tilde{D}(\Gamma_n(P, \Pi; -\xi), \Gamma_n(Q, \Pi; -\xi)) + \tilde{D}(\Gamma_n(Q, \Pi; -\xi), \Gamma_n(P, \Pi; -\xi))\}, \end{aligned}$$

where

$$\tilde{D}(u, v) = - \int_V u(\Delta v - \xi v) dV,$$

and where Π is the point of integration. (37) can be deduced in a similar way

to (12); see [14]. The last expression on the right of (37) vanishes on account of our choice of R . When

$$u(\Pi) = \gamma_n(P, \Pi; -\xi) + \gamma_n(Q, \Pi; -\xi)$$

and

$$u(\Pi) = \gamma_n(P, \Pi; -\xi) - \gamma_n(Q, \Pi; -\xi),$$

the expressions

$$E(u; \Gamma_n(P, \Pi; -\xi) + \Gamma_n(Q, \Pi; -\xi))$$

and

$$E(u; \Gamma_n(P, \Pi; -\xi) - \Gamma_n(Q, \Pi; -\xi))$$

attain their minimum values. These minimum values can be estimated as in Secs. 2—3 and we find (it is easily seen that we need only the estimation for the expression (14))

$$(38) \quad |\gamma_n(P, Q; -\xi) + \gamma_n(Q, P; -\xi)| \leq \text{const. } \xi^{\frac{N}{2}-2n-3}.$$

Interchanging P and Q in (36) and adding we obtain, on using (38)

$$(39) \quad (-1)^{p+1} \xi^{p+1} \sum_{m=0}^{\infty} \frac{\phi_m(P) \phi_m(Q)}{(\lambda_m + \xi) \lambda_m^{p+1}} = \sum_{p=0}^p A_p(P, Q) \xi^p + O(\xi^{\frac{N}{2}-2n-3}).$$

From this relation it follows as in Sec. 5 that the Dirichlet's series

$$(40) \quad \sum_{m=0}^{\infty} \frac{\phi_m(P) \phi_m(Q)}{\lambda_m^s}$$

which has a finite abscissa of convergence can be continued analytically to the left of this line and represents a regular function $\zeta(s, P, Q)$ for $\Re s > \frac{1}{2}N - 2n - 3$. $\zeta(s, P, Q) = 0$ when s is equal to a non-positive integer ($> \frac{1}{2}N - 2n - 3$).

When $\lambda_0 = 0$ we consider (40) but with summation from $m = 1$. Using instead of (36) the formula (see Sec. 4, formula (30))

$$\begin{aligned} \Gamma_n(P, Q; -\xi) - \gamma_n(P, Q; -\xi) + \sum_{p=0}^p B_p(P, Q) \xi^p - \frac{1}{V\xi} \\ = (-1)^{p+1} \xi^{p+1} \sum_{m=1}^{\infty} \frac{\phi_m(P) \phi_m(Q)}{(\lambda_m + \xi) \lambda_m^{p+1}}. \end{aligned}$$

We find that the analytic continuation of

$$(41) \quad \sum_{m=1}^{\infty} \frac{\phi_m(P) \phi_m(Q)}{\lambda_m^s}$$

has the same properties as the analytic continuation of (40) with the only exception that the function $\zeta(s, P, Q)$ represented by (41) is not zero for $s = 0$ but has the value $-\frac{1}{V}$.

7. The series $\sum_{m=1}^{\infty} \lambda_m^{-s}$ in the case of a closed manifold. We add a discussion of the series

$$\sum_{m=1}^{\infty} \lambda_m^{-s}$$

in the case of a closed manifold V . In this case R can be chosen independent of P and we can make use of the inequality (15) with a constant independent of R . Integrating (30) over V we have on account of this inequality

$$(42) \quad (-1)^{p+1} \xi^{p+1} \sum_{m=1}^{\infty} \frac{1}{(\lambda_m + \xi) \lambda_m^{p+1}} = \int_V M_n(\xi, P) dV \\ + \sum_{r=0}^p \xi^r \int_V A_r(P) dV - \frac{1}{\xi} + O(\xi^{\frac{N}{2}-n-2}).$$

Using the same method as in Sec. 5 we obtain the

THEOREM. If V is closed the Dirichlet's series

$$\sum_{m=1}^{\infty} \lambda_m^{-s}$$

can be continued analytically to the left of its abscissa of convergence and the function thus obtained can be written in the form

$$\frac{1}{(2\sqrt{\pi})^N} \sum_{r=0}^n \frac{\int_V U_r(P, P) dV}{\Gamma\left(\frac{N}{2} - r\right) \left(s - \frac{N}{2} + r\right)} + R_n(s) \text{ when } N \text{ is odd,}$$

and

$$\frac{1}{(2\sqrt{\pi})^N} \sum_{r=0}^{\frac{N}{2}-1} \frac{\int_V U_r(P, P) dV}{\Gamma\left(\frac{N}{2} - r\right) \left(s - \frac{N}{2} + r\right)} + R_n(s) \text{ when } N \text{ is even.}$$

In both cases $R_n(s)$ is regular in the half-plane $\Re s > \frac{1}{2}N - n - 2$. For $s = 0$ the value of the analytic continuation is -1 if N is odd, and

$$\frac{1}{(2\sqrt{\pi})^N} \int_V U_{\frac{N}{2}}(P, P) dV - 1$$

if N is even. For s equal to a negative integer the analytic continuation is zero when N is odd, and when N is even its value is

$$\frac{\Gamma(1-s)}{(2\sqrt{\pi})^N} \int_V U_{\frac{N}{2}-s}(P, P) dV.$$

Ikehara's theorem gives the asymptotic distribution of the eigenvalues

$$N(\lambda_m < T) \sim \frac{VT^{\frac{N}{2}}}{(2\sqrt{\pi})^N \Gamma\left(\frac{N}{2} + 1\right)}$$

where $N(\lambda_m < T)$ denotes the number of eigenvalues $< T$.

REFERENCES

- [1] T. Carleman, "Propriétés asymptotiques des fonctions fondamentales des membranes vibrantes," *Skand. Matem. Kongress* (1934).
- [2] T. Carleman, "Über die asymptotische Verteilung der Eigenwerte partieller Differentialgleichungen," *Berichte Verhandl. Akad. Leipzig*, vol. 88 (1936).
- [3] R. Courant and D. Hilbert, *Methoden der Mathematischen Physik I* (Berlin, 1924).
- [4] R. Courant and D. Hilbert, *Methoden der Mathematischen Physik II* (Berlin, 1937).
- [5] M. G. Giraud, "Généralisation des problèmes sur les opérations du type elliptique," *Bull. Sci. Math.*, Series 2, vol. 56 (1932).
- [6] J. Hadamard, *Le problème de Cauchy et les équations aux dérivées partielles linéaires hyperboliques* (Paris, 1932).
- [7] S. Minakshisundaram, "A Generalization of Epstein Zeta Functions," will appear in this Journal.
- [8] S. Minakshisundaram, "Zeta-functions on the Sphere," will appear in *J. Indian Math. Soc.*
- [9] Å. Pleijel, "Propriétés asymptotiques des fonctions et valeurs propres de certains problèmes de vibrations," *Arkiv Mat., Astr. o. Fys.*, vol. 27 A, no. 13 (1940).
- [10] Å. Pleijel, "Asymptotic Properties of the Eigenfunctions of Certain Boundary-Value Problems of Polar Type," will appear in *Amer. J. Math.*
- [11] G. de Rham, "Sur la théorie des formes différentielles harmoniques," *Ann. Univ. Grenoble, Sec. Sci. Math. Phys.*, (N.S.) vol. 22 (1946).
- [12] G. M. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge, 1944).
- [13] H. Weyl, "Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung)," *Math. Ann.*, vol. 71 (1911).
- [14] N. Wiener, "Tauberian Theorems," *Ann. of Math.*, vol. 33 (1932).

Andhra University
Waltair, South India

ON THE MOTION OF THREE VORTICES

J. L. SYNGE

1. Introduction. In a perfect incompressible fluid extending to infinity, the determination of the motion of N parallel rectilinear vortex filaments involves the solution of N non-linear differential equations, each of the first order. The method of Kirchhoff¹ provides certain constants of the motion. If we describe the positions of the vortices by their point-traces on a plane perpendicular to them, the following facts follow from the theory of Kirchhoff:

(1.1) The mean centre of the system is fixed.

$$(1.2) \quad \sum'_{m,n} \kappa_m \kappa_n \log r_{mn} = \text{const.}$$

$$(1.3) \quad \sum_m \kappa_m r_m^2 = \text{const.}$$

Here the summations cover the range $1, 2, \dots, N$; the prime indicates that $m = n$ is omitted; κ_m are the strengths of the vortices; r_{mn} is the distance between the vortices of strengths κ_m and κ_n ; r_m is the distance of the vortex of strength κ_m from a fixed point.

In this paper we shall be concerned solely with the configurations of the vortex system, understanding by *configuration* the geometrical figure formed by the vortices, without regard to rigid body displacements of that figure. Thus, if a system of three vortices forms a triangle with sides of fixed lengths throughout the motion, we say that the configuration is fixed.

The following theorems, applicable to a system consisting of any number of vortices, are obvious from the usual equations of vortex motion, and are quoted here for reference.

THEOREM 1: *If, given a configuration, the strengths of all the vortices are suddenly reversed, the system retraces the sequence of configurations through which it has come.*

THEOREM 2: *Given at $t = t_0$ a configuration in which all the vortices are collinear, then the configurations at times $t = t_0 \pm \tau$ are reflections of one another for all values of τ .*

THEOREM 3: *A system cannot pass through more than two distinct collinear configurations; the times required to pass from one collinear configuration to the other are all the same.*

THEOREM 4: *Suppose that there are two systems of vortices, S_1 and S_2 , each consisting of the same number of vortices, and the strengths of the vortices in S_1*

Received May 2, 1948.

¹Cf. Sir H. Lamb, *Hydrodynamics* (Cambridge, 1932), 230; H. Villat, *Leçons sur la théorie des tourbillons* (Paris, 1930), 46.

being those of S_1 all multiplied by the same factor K^2 ; suppose further that initially the configurations are similar, without reflection, the lengths in S_2 being those in S_1 all multiplied by the factor L . Then the subsequent configuration of S_2 after time t_2 is similar, without reflection, to the configuration of S_1 after time t_1 , where $t_2 = t_1(L^2/K^2)$.

As an immediate consequence of (1.2) and (1.3), we have the following result:

THEOREM 5: *If the strengths of all the vortices have the same sign, their mutual distances are bounded above and below for all time, positive and negative.*

No further general results appear to be available, so we turn to special cases. The general case can be specialized in a number of ways. We might specialize the strengths of vortices, perhaps choosing them all of the same strength, or plus and minus one fixed value. On the other hand we might specialize by restricting the number of vortices in the system, and this is in fact the specialization we shall adopt.

Since the case of two vortices is trivial, we turn to the case of three vortices, without imposing any particular *a priori* condition on their strengths. This is precisely the problem discussed by W. Gröbli² over seventy years ago. However, he was interested in obtaining formal analytic solutions for the motion, and found it necessary at an early stage to specialize the strengths of the vortices. He seems to have missed the interesting fact that the motions may be classified according to the positive or negative character of the sum of the products of the strengths in pairs, $\kappa_2\kappa_3 + \kappa_3\kappa_1 + \kappa_1\kappa_2$. It seems appropriate therefore to take up this problem again, concentrating on a qualitative classification of all possible motions rather than on the development of analytic solutions. The basic equations (2.5) are the same as those of Gröbli, but are obtained here in a simpler way. The representation of the motions by trilinear coordinates is believed to be new.

2. The equations of motion and their integrals. Let $\kappa_1, \kappa_2, \kappa_3$ be the strengths of the three vortices (i.e. the circulations around them), and R_1, R_2, R_3 the lengths of the sides of the triangle formed by them, R_1 being opposite κ_1 , and so on, so that, in the notation of (1.2), $R_1 = r_{23}$, etc. In accordance with the usual convention, we regard a strength as positive when it gives a counter-clockwise circulation. It is convenient to get rid of the factor 2π by defining

$$(2.1) \quad k_1 = \kappa_1/2\pi, \quad k_2 = \kappa_2/2\pi, \quad k_3 = \kappa_3/2\pi.$$

It is assumed that none of the three strengths vanishes.

Consider the rate of increase $R'_1 = dR_1/dt$ of the side R_1 . The motions due to the vortices k_2 and k_3 at its extremities contribute nothing to R'_1 . One end of R_1 , viz. k_2 , has due to k_1 a velocity of magnitude k_1/R_3 perpendicular to R_3 , and the other end, viz. k_3 , has due to k_1 a velocity of magnitude k_1/R_2 per-

²*Vierteljahrsschrift der naturforschenden Gesellschaft in Zürich*, vol. 22 (1877), 37-81, 120-167. Gröbli also investigated certain cases of symmetry for N vortices.

pendicular to R_2 . Let $\theta_1, \theta_2, \theta_3$ be the angles of the triangle formed by the vortices. Then, on reference to Figure 1, it is seen that

$$(2.2) \quad R'_1 = \epsilon k_1 (R_2^{-1} \sin \theta_3 - R_3^{-1} \sin \theta_2),$$

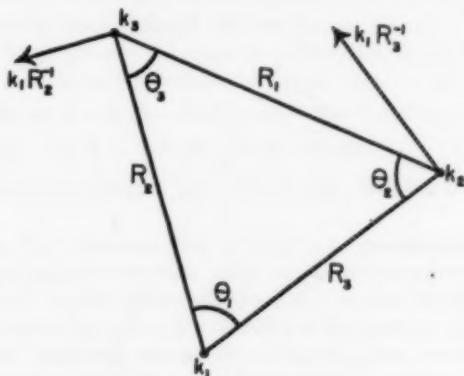


FIG. 1

Rate of growth of a side of the triangle.

where $\epsilon = +1$ or -1 according as the circuit of the triangle in the order $k_1 k_2 k_3$ is positive or negative respectively (counter-clockwise or clockwise). Let A denote the area of the triangle, prefixed by a plus or minus sign according as the above circuit is positive or negative. Then ϵA is positive, and

$$(2.3) \quad \epsilon A = \frac{1}{2} R_2 R_3 \sin \theta_1 = \frac{1}{2} R_3 R_1 \sin \theta_2 = \frac{1}{2} R_1 R_2 \sin \theta_3.$$

We have also the formula

$$(2.4) \quad \epsilon A = [s(s - R_1)(s - R_2)(s - R_3)]^{1/2},$$

$$s = \frac{1}{2}(R_1 + R_2 + R_3).$$

If we substitute from (2.3) in (2.2) and the two similar equations, we get

$$(2.5) \quad \begin{aligned} k_1^{-1} R_1 R'_1 &= 2A(R_2^{-2} - R_3^{-2}), \\ k_2^{-1} R_2 R'_2 &= 2A(R_3^{-2} - R_1^{-2}), \\ k_3^{-1} R_3 R'_3 &= 2A(R_1^{-2} - R_2^{-2}). \end{aligned}$$

Adding and integrating, we get

$$(2.6) \quad k_1^{-1} R_1^2 + k_2^{-1} R_2^2 + k_3^{-1} R_3^2 = a,$$

where a is a constant. If we multiply (2.5) in order by $R_1^{-2}, R_2^{-2}, R_3^{-2}$, add, and integrate, we get

$$(2.7) \quad k_1^{-1} \log R_1 + k_2^{-1} \log R_2 + k_3^{-1} \log R_3 = b,$$

where b is a constant. This is the same as Kirchhoff's equation (1.2), and (2.6) is equivalent to (1.3), but more convenient for our purpose because

expressed in terms of the sides of the triangle. The above equations were given by Gröbli (*loc. cit.*).

The differential equations (2.5), with their integrals (2.6) and (2.7), form the basis of our work. To these we shall add another equation, obtained by differentiating (2.4) and then substituting for R'_1, R'_2, R'_3 . In this way we get

$$(2.8) \quad A' = f(R_1, R_2, R_3),$$

where

$$(2.9) \quad f(R_1, R_2, R_3) \\ = \frac{1}{2}[\Sigma k_1 R_1^{-1}(R_2^{-2} - R_3^{-2})][(s - R_1)(s - R_2)(s - R_3) + s\Sigma(s - R_2)(s - R_3)] \\ - s\Sigma k_1 R_1^{-1}(R_2^{-2} - R_3^{-2})(s - R_2)(s - R_3).$$

Here and later, Σ indicates summation over a cyclic permutation of suffixes.

3. Fixed configurations. Let us now seek necessary and sufficient conditions that the configuration of the three vortices remains fixed, so that the motion is a rigid body motion. If the configuration is fixed, then $R'_1 = R'_2 = R'_3 = 0$ and so by (2.5) we must have either $R_1 = R_2 = R_3$ (equilateral configuration), or $A = 0$ (collinear configuration). These are necessary conditions. Any equilateral configuration does remain fixed, as was pointed out by Gröbli (*loc. cit.*), and this is a sufficient condition. But $A = 0$ is not a sufficient condition for fixity. At first sight this appears to be in conflict with (2.5). Suppose we take for R_1, R_2, R_3 any three constant values satisfying one of the equations

$$(3.1) \quad R_1 = R_2 + R_3, \quad R_2 = R_3 + R_1, \quad R_3 = R_1 + R_2,$$

such values make $A = 0$ by (2.4), and hence these values constitute a formal solution of (2.5). However, it is a singular solution, and does not in general satisfy the full set of equations of vortex motion. In order that the collinear configuration may remain fixed, it is further necessary that $A' = 0$, or

$$(3.2) \quad f(R_1, R_2, R_3) = 0,$$

where f is as in (2.9). We may sum up as follows:

THEOREM 6: *Necessary and sufficient conditions for a fixed configuration are either that the initial configuration be equilateral, or that it be collinear, satisfying (3.2).*

4. Variable configurations and the trilinear representation. The values of R_1, R_2, R_3 determine a configuration to within a reflection. Thus we might discuss changes in configuration by following a representative point in a space in which R_1, R_2, R_3 are taken as rectangular Cartesian coordinates. Since these quantities are essentially positive, we would be concerned only with the positive octant. Collision of the representative point with one of the walls of this octant would correspond to a collision of two of the vortices. The motion of the system would correspond to a curve of intersection of surfaces (2.6) and (2.7), the sense in which the curve is described being determined by reference to (2.5), with use of the fact that t increases. But the representative point is

further restricted since R_1, R_2, R_3 must always satisfy the triangle inequalities (4.1)

$$R_1 \leq R_2 + R_3, R_2 \leq R_3 + R_1, R_3 \leq R_1 + R_2.$$

In fact, the planes (3.1) form boundaries in the representative space which the representative point is forbidden to cross. If the representative point meets one of the planes (3.1), the configuration becomes collinear. Then, by Theorem 2, the system passes back through the same sequence of configurations but with the orientation reversed; the representative point moves back along the curve by which it came to the collinear configuration.

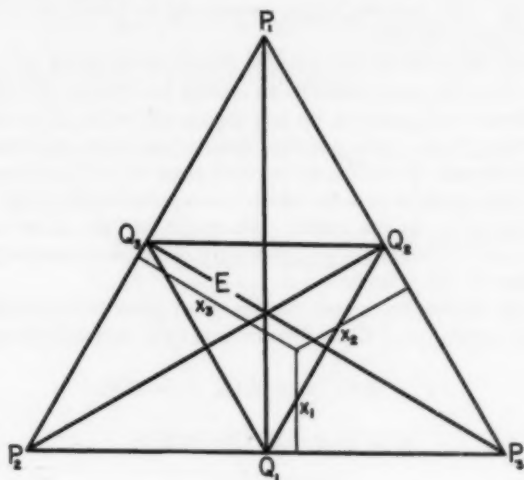


FIG. 2

The trilinear representation.

However, there is another and better representation by trilinear coordinates in a plane, as shown in Figure 2, and that is the representation which will be used in this paper. $P_1P_2P_3$ is an equilateral triangle of unit height, and x_1, x_2, x_3 are trilinear coordinates, i.e. the distances of a general point from the sides of the triangle $P_1P_2P_3$; these values satisfy

$$(4.2) \quad x_1 + x_2 + x_3 = 1.$$

Now put

$$(4.3) \quad \begin{aligned} x_1 &= R_1(R_1 + R_2 + R_3)^{-1}, \\ x_2 &= R_2(R_1 + R_2 + R_3)^{-1}, \\ x_3 &= R_3(R_1 + R_2 + R_3)^{-1}, \end{aligned}$$

and so connect the points of the representative plane with the configurations of the vortex system. To each configuration of the system there corresponds a unique x -point, with one exception: a triple collision ($R_1 = R_2 = R_3 = 0$) is

not represented. On the other hand, to a given x -point there corresponds a single infinity of configurations, all similar to one another, together with the reflections of those configurations. The centroid E of the triangle ($x_1 = x_2 = x_3 = 1/3$) corresponds to all equilateral configurations.

Let $Q_1Q_2Q_3$ be the middle points of the sides of the triangle $P_1P_2P_3$. On Q_2Q_3 we have $x_1 = \frac{1}{2}$ and hence $x_1 = x_2 + x_3$ or $R_1 = R_2 + R_3$. In fact, Q_2Q_3 corresponds to the first of (3.1), and the three sides of the triangle $Q_1Q_2Q_3$ correspond to the three planes (3.1) which the representative point is forbidden to cross. Since E is certainly permitted, the permitted region is the interior of the triangle $Q_1Q_2Q_3$. The points $Q_1Q_2Q_3$ correspond to collisions of the vortices, k_2 and k_3 colliding at Q_1 , etc.

All points on the sides of the triangle $Q_1Q_2Q_3$ correspond to collinear configurations. Since the configuration can change its orientation only by passing through a collinear configuration, we may use the two sides of the representative plane, all configurations with positive orientation being represented on the front of the plane and all configurations with negative orientation on the back. The sides $Q_1Q_2Q_3$ are then cuts by which the representative point passes from one side of the plane to the other. We might in fact throw away all the diagram except the triangle $Q_1Q_2Q_3$, and allow the representative point to pass round the edges of this triangle.

As the system moves, the representative point describes a curve C . To find the differential equations of C , we differentiate (4.3) and substitute from (2.5). This gives

$$(4.4) \quad x'_1 = KH_1, \quad x'_2 = KH_2, \quad x'_3 = KH_3,$$

where

$$(4.5) \quad K = 2AR_1^{-2}R_2^{-2}R_3^{-2}(\Sigma R_1)^2,$$

and

$$(4.6) \quad \begin{aligned} H_1 &= -k_1x_1(x_2^2 - x_3^2) + x_1\Sigma k_1x_1(x_2^2 - x_3^2), \\ H_2 &= -k_2x_2(x_3^2 - x_1^2) + x_2\Sigma k_1x_1(x_3^2 - x_1^2), \\ H_3 &= -k_3x_3(x_1^2 - x_2^2) + x_3\Sigma k_1x_1(x_1^2 - x_2^2). \end{aligned}$$

We check that $H_1 + H_2 + H_3 = 0$, as of course it must be, by (4.2).

By (4.4.) we have

$$(4.7) \quad \frac{dx_1}{H_1} = \frac{dx_2}{H_2} = \frac{dx_3}{H_3} = Kdt.$$

The first two of these equations define a congruence of x -curves, and this congruence defines the behaviour of the configuration, except for orientation, rate of change, and scale. However, orientation is determined by the side of the representative plane on which the point lies, and rate of change is given by (4.4). As regards scale, if the shape of the configuration is given, its size may in general be determined by (2.6) or (2.7), the values of the constants a and b being given by the initial configuration. There is, however, one exceptional case, and this we shall now discuss.

The integrals (2.6) and (2.7) may be written

$$(4.8) \quad k_1^{-1}x_1^2 + k_2^{-1}x_2^2 + k_3^{-1}x_3^2 = a(R_1 + R_2 + R_3)^{-2},$$

$$(4.9) \quad k_1^{-1} \log x_1 + k_2^{-1} \log x_2 + k_3^{-1} \log x_3 = b - (k_1^{-1} + k_2^{-1} + k_3^{-1}) \log (R_1 + R_2 + R_3).$$

If $a = 0$ and

$$(4.10) \quad k_2k_3 + k_3k_1 + k_1k_2 = 0,$$

then $(R_1 + R_2 + R_3)$ disappears from (4.8) and (4.9). In this exceptional case, the values of x_1, x_2, x_3, a, b fail to determine the values of R_1, R_2, R_3 . We may state the following results.

THEOREM 7: *If the strengths of the vortices do not satisfy (4.10), and b is known from an initial configuration, then to each x -point there corresponds by (4.9) a unique configuration, except for orientation.*

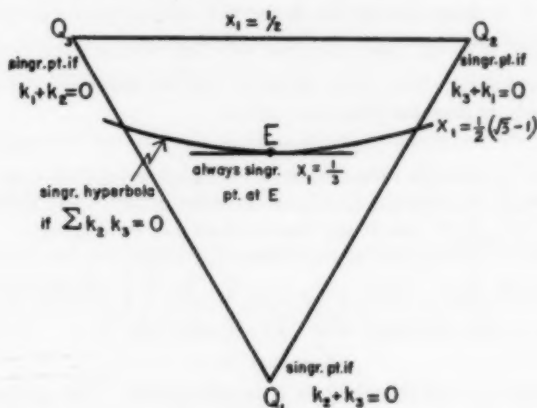


FIG. 3

Singular points.

(Hyperbola drawn for $2k_1 = -k_2 = -k_3$.)

THEOREM 8: *If the strengths of the vortices satisfy (4.10), then to each x -point on the conic*

$$(4.11) \quad k_2k_3x_1^2 + k_3k_1x_2^2 + k_1k_2x_3^2 = 0$$

there corresponds a single infinity of similar configurations of both orientations; to each x -point lying off the conic (4.11) there corresponds by (4.8) a unique configuration, except for orientation.

It is easily seen that, under the condition (4.10), the conic (4.11) is a hyperbola. It passes through the centroid E , and meets two sides of the triangle $Q_1Q_2Q_3$, each in one point. At E the tangent to (4.11) has the direction given by

$$(4.12) \quad dx_1 : dx_2 : dx_3 = k_1(k_2 - k_3) : k_2(k_3 - k_1) : k_3(k_1 - k_2).$$

The hyperbola is shown in Figure 3 for the case

$$(4.13) \quad 2k_1 = -k_2 = -k_3.$$

It is important to know that *no curve C can cut a median of the triangle $Q_1Q_2Q_3$ in an infinite number of points.* To show this, we consider the median P_1Q_1 , on which we have

$$x_2 = x_3 = \frac{1}{2}(1 - x_1).$$

By (4.8) and (4.9) we have at an intersection of a curve C with the median P_1Q_1

$$(4.14) \quad k_1^{-1}x_1^2 + (k_2^{-1} + k_3^{-1}) \frac{1}{4}(1 - x_1)^2 = a/4s^2,$$

$$k_1^{-1} \log x_1 + (k_2^{-1} + k_3^{-1}) \log \frac{1}{2}(1 - x_1) = b - (k_1^{-1} + k_2^{-1} + k_3^{-1}) \log 2s,$$

where, as earlier, $2s = R_1 + R_2 + R_3$. If we eliminate s , we get an equation in x_1 , a , b ; for a given curve C the constants a and b are assigned, and this equation determines the values of x_1 corresponding to the intersections of C with P_1Q_1 . It is clear that in the range $0 \leq x_1 \leq 1$ there can be at most a finite number of solutions, and so the result is proved.

5. Singular points. The most powerful way of studying the congruence (4.7) is through its singular points, at which

$$(5.1) \quad H_1 = H_2 = H_3 = 0.$$

On account of the triangle inequalities (4.1), we are interested only in singular points lying inside the triangle $Q_1Q_2Q_3$ or on its boundary. Let us first examine the points Q_1 , Q_2 , Q_3 , to see if any one of them can be singular.

At Q_1 we have $x_1 = 0$, $x_2 = x_3 = \frac{1}{2}$; hence, by (4.6),

$$(5.2) \quad H_1 = 0, H_2 = -\frac{1}{8}k_2 + \frac{1}{16}(k_2 - k_3), H_3 = \frac{1}{8}k_3 + \frac{1}{16}(k_2 - k_3).$$

These equations are consistent with (5.1) if, and only if,

$$(5.3) \quad k_2 + k_3 = 0.$$

When this condition is satisfied, Q_1 is a singular point. The points Q_2 and Q_3 may of course be discussed in exactly the same way.

For all points in the triangle $Q_1Q_2Q_3$ or on its boundary, other than the vertices Q_1 , Q_2 , Q_3 , we have x_1, x_2, x_3 all different from zero. Then, if we substitute in (5.1) from (4.6), we can divide across by these factors, and obtain

$$(5.4) \quad x_2^2 - x_3^2 = k_1^{-1}\theta, \quad x_3^2 - x_1^2 = k_2^{-1}\theta, \quad x_1^2 - x_2^2 = k_3^{-1}\theta,$$

$$\theta = \Sigma k_1 x_1 (x_2^2 - x_3^2).$$

Addition gives

$$(5.5) \quad \theta \Sigma k_1^{-1} = 0.$$

Suppose first that $\theta = 0$; then (5.4) give $x_1 = x_2 = x_3 = 1/3$. Thus the point E is a singular point, as is indeed obvious. On the other hand, if (4.10) is satisfied, then (5.5) is satisfied with $\theta \neq 0$. If we multiply (5.4) in order by x_1^2, x_2^2, x_3^2 and add, we get

$$(5.6) \quad \Sigma k_1^{-1} x_1^2 = 0$$

which is the same equation as (4.11). All singular points (other than Q_1, Q_2, Q_3 , discussed above) must lie on this conic. Moreover it is easy to see that,

if (4.10) is satisfied, then every point on the conic (4.11) or (5.6) is a singular point. We have already remarked that this conic is a hyperbola.

Let us sum up our conclusions about singular points as follows.

THEOREM 9: *The singular points of the congruence (4.7), inside or on the triangle $Q_1Q_2Q_3$, are as follows. If*

$$(5.7) \quad k_2k_3 + k_3k_1 + k_1k_2 \neq 0,$$

and

$$(5.8) \quad k_2 + k_3 \neq 0, k_3 + k_1 \neq 0, k_1 + k_2 \neq 0,$$

then the only singular point is at E (equilateral configuration). If

$$(5.9) \quad k_2k_3 + k_3k_1 + k_1k_2 = 0,$$

then (5.8) are necessarily true; the singular points make up the hyperbola (4.11), which passes through E . If

$$(5.10) \quad k_2 + k_3 = 0, k_3 + k_1 \neq 0, k_1 + k_2 \neq 0,$$

then (5.7) is necessarily true; the only singular points are at E and Q_1 . Similar results hold on permuting suffixes in (5.10). If

$$(5.11) \quad k_1 = -k_2 = -k_3,$$

the only singular points are at E, Q_2, Q_3 . Similar results hold on permutation of suffixes.

These results are shown in Figure 3.

6. Behaviour of representative curves near the point E . To explore the curves near the point E , we put

$$(6.1) \quad x_1 = y_1 + 1/3, x_2 = y_2 + 1/3, x_3 = y_3 + 1/3,$$

so that

$$(6.2) \quad y_1 + y_2 + y_3 = 0.$$

Then (4.6) gives, to the first order in y_1, y_2, y_3 ,

$$(6.3) \quad \begin{aligned} H_1 &= -\frac{2}{3} k_1(y_2 - y_3) + \frac{2}{3^2} \Sigma k_1(y_2 - y_3), \\ H_2 &= -\frac{2}{3} k_2(y_3 - y_1) + \frac{2}{3^2} \Sigma k_1(y_2 - y_3), \\ H_3 &= -\frac{2}{3} k_3(y_1 - y_2) + \frac{2}{3^2} \Sigma k_1(y_2 - y_3). \end{aligned}$$

As in (4.7) we have, as differential equations of the congruence,

$$(6.4) \quad \frac{dy_1}{H_1} = \frac{dy_2}{H_2} = \frac{dy_3}{H_3} = K dt.$$

It is convenient to define

$$(6.5) \quad z_1 = y_2 - y_3, z_2 = y_3 - y_1, z_3 = y_1 - y_2,$$

so that, by (6.2),

$$(6.6) \quad y_1 = -\frac{1}{3}(z_2 - z_3), y_2 = -\frac{1}{3}(z_3 - z_1), y_3 = -\frac{1}{3}(z_1 - z_2).$$

From (6.4) we obtain

$$(6.7) \quad \frac{dz_1}{L_1} = \frac{dz_2}{L_2} = \frac{dz_3}{L_3},$$

where

$$\begin{aligned}
 (6.8) \quad L_1 &= -\frac{3}{2} (H_2 - H_3) = k_2 z_2 - k_3 z_3, \\
 L_2 &= -\frac{3}{2} (H_3 - H_1) = k_3 z_3 - k_1 z_1, \\
 L_3 &= -\frac{3}{2} (H_1 - H_2) = k_1 z_1 - k_2 z_2.
 \end{aligned}$$

If we put each fraction in (6.7) equal to ds , we have the equations

$$\begin{aligned}
 (6.9) \quad \frac{dz_1}{ds} &= k_2 z_2 - k_3 z_3, \\
 \frac{dz_2}{ds} &= -k_1 z_1 + k_3 z_3, \\
 \frac{dz_3}{ds} &= k_1 z_1 - k_2 z_2.
 \end{aligned}$$

We have, by (6.5),

$$(6.10) \quad z_1 + z_2 + z_3 = 0,$$

and so the first two of (6.9) give

$$\begin{aligned}
 (6.11) \quad \frac{dz_1}{ds} &= k_3 z_1 + (k_2 + k_3) z_2, \\
 \frac{dz_2}{ds} &= -(k_1 + k_3) z_1 - k_2 z_2.
 \end{aligned}$$

The solutions are of the form $\exp(\lambda s)$, where the eigenvalues λ satisfy

$$(6.12) \quad \begin{vmatrix} k_3 - \lambda & k_2 + k_3 \\ -k_1 - k_3 & -k_2 - \lambda \end{vmatrix} = 0,$$

or

$$(6.13) \quad \lambda^2 = -\Sigma k_2 k_3.$$

Three cases arise:

Case I: $\Sigma k_2 k_3 > 0$; eigenvalues pure imaginary;

Case II: $\Sigma k_2 k_3 < 0$; eigenvalues real, one positive and one negative;

Case III: $\Sigma k_2 k_3 = 0$; eigenvalues both zero.

7. Case I: $k_2 k_3 + k_3 k_1 + k_1 k_2 > 0$.

In Case I the curves (6.9) are closed curves, surrounding the point E . However, (6.9) is only a linear approximation to the curves C , and it does not follow immediately that the curves C are closed. But if a curve C is not closed, then, since it cannot intersect itself, it must cut a median $P_1 Q_1$ in an infinite number of points. This we have shown earlier to be impossible. Hence all curves C near E are in fact closed curves (Figure 4). The sense in which such a curve is described depends on the initial orientation of the triangle (cf. (4.4), (4.5)).

If we expand the orbit (which, roughly speaking, means bringing two of the vortices closer together, since Q_1, Q_2, Q_3 correspond to collisions), we shall reach an orbit C_0 which touches the periphery $Q_1Q_2Q_3$ at a point corresponding to a fixed collinear configuration. This configuration will be approached as a limit, not attained in finite time.

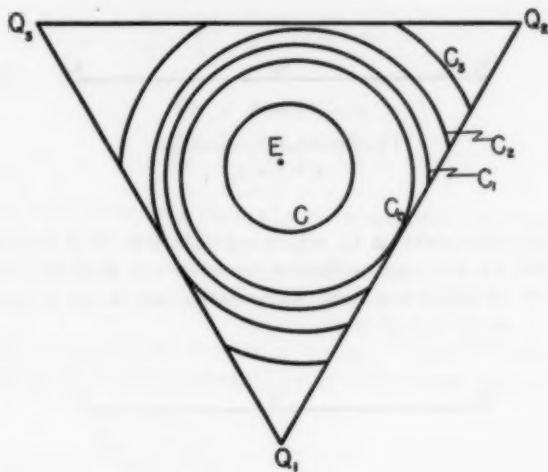


FIG. 4

Representative curves for Case I: $\Sigma k_i k_j > 0$.

It is interesting to consider here the particular case, $k_1 = k_2 = k_3$, which of course belongs to Case I. Now the figure is symmetric, and C_0 will touch all three sides of $Q_1Q_2Q_3$. Thus the system, if started on such a curve, will oscillate in infinite time between two fixed collinear configurations, these two configurations being different. For three equal vortices, the only fixed collinear configurations are those in which the vortices are equally spaced (Figure 5). Such a configuration, if slightly disturbed, will pass in a long time near to one of the configurations shown in Figure 6. Equation (2.5) tells us the lengths in Figure 6 are the same as those in Figure 5. If the representative curve of the disturbed motion does not meet $Q_1Q_2Q_3$ (i.e. if it belongs to the class C of Figure 4), then all three configurations of Figures 5 and 6 will be approached one after another. By symmetry, the representative curve cannot belong to class C_1 or class C_2 . If it is of class C_3 , then the motion is an oscillation between a collinear configuration adjacent to that shown in Figure 5 and a collinear configuration adjacent to one of those shown in Figure 6. These oscillations between configurations which differ only through interchange of vortices of equal strength appear rather interesting.

In the general case of unequal strengths, contact will be established first with one side of $Q_1Q_2Q_3$, as for C_0 in Figure 4. When we expand the orbit further to C_1 , we get an oscillation, performed in finite time, between two collinear configurations which are actually the same configuration. We may think of the return journey as performed on the back of the representative plane; it has reversed orientation.



FIG. 5

Fixed collinear configuration.

$$(k_1 = k_2 = k_3)$$

Further expansion gives us C_2 , which cuts one side of $Q_1Q_2Q_3$ and touches another. Here we have an oscillation between two different collinear configurations, one of which is a fixed configuration and is not attained in finite time.

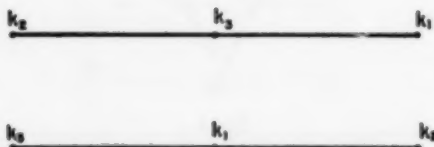


FIG. 6

Transforms of configuration of FIG. 5.

The final stage is C_3 , representing an oscillation in finite time between two different collinear configurations.

This exhausts the possibilities in Case I. In this case the equilateral configuration is of course stable for small disturbances.

8. Case II: $k_2k_3 + k_3k_1 + k_1k_2 < 0$.

Here the eigenvalues are $\pm\mu$, where

$$(8.1) \quad \mu = (-\Sigma k_2k_3)^{\frac{1}{2}} > 0.$$

The solutions of (6.11) are

$$(8.2) \quad \begin{aligned} z_1 &= A_1 e^{\mu z} + B_1 e^{-\mu z}, \\ z_2 &= A_2 e^{\mu z} + B_2 e^{-\mu z}, \end{aligned}$$

where

$$(8.3) \quad \begin{aligned} A_1(\mu - k_3) - A_2(k_2 + k_3) &= 0, \\ B_1(-\mu - k_3) - B_2(k_2 + k_3) &= 0. \end{aligned}$$

As $s \rightarrow \infty$, the curve recedes asymptotically in the direction

$$(8.4) \quad z_1/z_2 = A_1/A_2 = (k_2 + k_3)/(\mu - k_3),$$

and as $s \rightarrow -\infty$, we have a curve coming in asymptotically from the direction

$$(8.5) \quad z_1/z_2 = B_1/B_2 = -(k_2 + k_3)/(\mu + k_3).$$

These directions may be expressed symmetrically. They correspond to values of z_1, z_2, z_3 which make

$$\frac{dz_1}{ds} : \frac{dz_2}{ds} : \frac{dz_3}{ds} = z_1 : z_2 : z_3,$$

and so, by (6.7), they satisfy

$$(8.6) \quad \begin{aligned} \lambda z_1 - k_2 z_2 + k_3 z_3 &= 0, \\ k_1 z_1 + \lambda z_2 - k_3 z_3 &= 0, \\ -k_1 z_1 + k_2 z_2 + \lambda z_3 &= 0, \\ z_1 + z_2 + z_3 &= 0. \end{aligned}$$

If we multiply the first three of these equations in order by k_1, k_2, k_3 , and add, and then solve with the last of (8.6), we get

$$(8.7) \quad \begin{aligned} z_1 : z_2 : z_3 &= \lambda(k_2 - k_3) + 3k_2k_3 - \Sigma k_2k_3 \\ &\quad : \lambda(k_3 - k_1) + 3k_3k_1 - \Sigma k_3k_1 \\ &\quad : \lambda(k_1 - k_2) + 3k_1k_2 - \Sigma k_1k_2. \end{aligned}$$

We are to put $\lambda = \pm \mu$ to get the two directions. Figure 7 shows such directions (D_1, D_2, D_3, D_4) and the general nature of the curves near E .

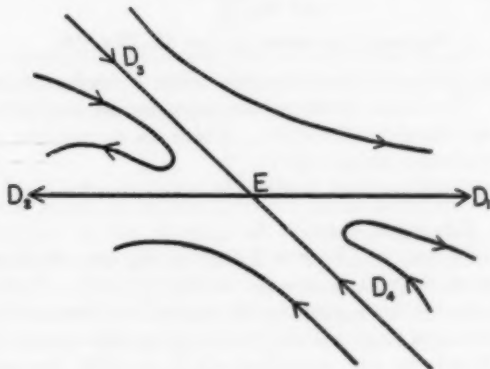


FIG. 7

Representative curves near E .
Case II: $\Sigma k_2k_3 < 0$.

The curves which start from E in the directions D_1, D_2, D_3, D_4 must pass out across the periphery $Q_1Q_2Q_3$ since they cannot cross nor can they cut a median of the triangle an infinite number of times. Similarly all representative curves must cross the periphery $Q_1Q_2Q_3$. The general nature of the pattern is shown in Figure 8.

The curves labelled D_1, D_2, D_3, D_4 represent motions in which the configuration oscillates between the equilateral configuration and a collinear configuration. The time of approach to E , or recession from it, is infinite. The other

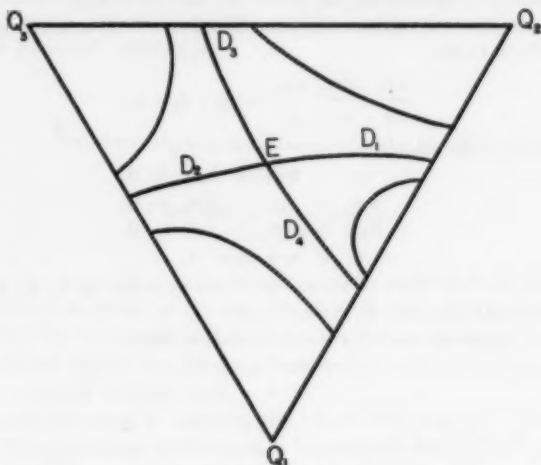


FIG. 8

Representative curves for Case II: $\sum k_i k_j < 0$.

curves represent oscillations between two collinear configurations, not necessarily distinct. The times involved are finite unless the collinear configuration involved is a fixed configuration. There are no periodic motions which do not include collinear configurations.

The equilateral configuration is unstable in this case for small disturbances.

9. Case III: $k_2 k_3 + k_3 k_1 + k_1 k_2 = 0$.

We have already seen in Theorem 8 that in this case there is a hyperbola (4.11) composed of singular points ($H_1 = H_2 = H_3 = 0$). If the initial configuration is represented by a point on this hyperbola, then by (4.4) the representative point remains fixed. Thus the configuration remains fixed in *shape*. To see how it changes its size, we refer to (2.5), in which the right-hand sides are now constants. It is clear that the squares of the sides increase or decrease linearly with time, remaining fixed in length only if the representative point is at E .

If initially the representative point does not lie on the hyperbola (4.11), then both shape and size change. This hyperbola forms a barrier which the representative point cannot cross. Hence the motion consists of an oscillation between collinear configurations.

*Institute for Advanced Studies,
Dublin, Eire*

ON SURFACE WAVES

ALEXANDER WEINSTEIN

1. Introduction. The linearized theory of surface waves leads to several mixed boundary value problems which have been investigated by various methods. As the physical background of the theory has been repeatedly discussed, it will suffice to deal here mainly with the mathematical aspect of the question.

Let D be a finite or infinite domain in the (x, y) -plane and let $\phi(x, y)$ denote a function in D satisfying one of the following differential equations

$$(1.1) \quad \frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} = 0,$$

$$(1.2) \quad \frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} - k^2 \phi = 0$$

where k^2 denotes a positive constant. Let n denote the external normal to the boundary C of D . The boundary condition on the part of C corresponding to the free surface of the fluid is given by the equation

$$(1.3) \quad \frac{\partial \phi}{\partial n} = p\phi,$$

The *positive* constant p is in some cases an unknown parameter.

The boundary condition on the part of C corresponding to the rigid part of the boundary is given by the equation

$$(1.4) \quad \frac{\partial \phi}{\partial n} = 0.$$

In the classical theory ϕ denotes (up to a factor depending on the time t) the velocity potential in the physical (x, y) -plane. However, in Levi-Civita's theory of plane waves, the independent variables are the velocity potential and the stream function. The unknown function ϕ denotes in this case the angle which the velocity makes with the horizontal direction. The condition (1.3) remains unchanged in form, but (1.4) has to be replaced by the condition

$$(1.5) \quad \phi = 0.$$

In most cases in application the domain D extends to infinity and is bounded by straight lines. However in Levi-Civita's theory of periodic waves, D can be mapped conformally on a finite domain such as a circle or circular ring without changing the form of the boundary conditions.

It should be emphasized that (1.3) differs essentially from the boundary condition discussed by Fourier in his classical theory of heat conduction. In Fourier's case p is essentially negative, a fact which implies that, for a finite

domain D , the corresponding boundary value problem admits only the trivial solution $\phi = 0$. The situation is, however, different in the case of surface waves where the corresponding boundary value problem admits one or even several non-trivial solutions for certain positive values of the constant p . The important question of the uniqueness of the solution has been overlooked by standard treatises on hydrodynamics. Besides its intrinsic mathematical interest, a survey of all solutions is of great importance for the following reasons: first, the solutions of the linearized problem give a first approximation to the exact non-linear theory of a surface wave; second, the superposition of two standing waves obtained from two different solutions of the linearized problem leads to a travelling wave as required by the theory.

There are at present three methods of approach to the various boundary problems encountered in the theory of surface waves:

- (i) The eigenvalue method.
- (ii) The method of reduction.
- (iii) The method of singular integral equations.

Some of the problem can at present be discussed only by the first or by the third method. However this is not the case for problems discussed up to now by the second method alone. It is the purpose of the present paper to show that a combination of the method of reduction with the eigenvalue method leads to more complete results than the application of the reduction method alone.

2. The eigenvalue method. This method has been developed by A. Weinstein [3] in connection with a problem in Levi-Civita's theory. For modification of this method see the papers by G. Hoheisel [4], S. Bochner [5], J. L. B. Cooper [6] and A. E. Heins [7]. As an illustration we shall use this method for the complete solution of *Airy's Problem* which corresponds to the hydrodynamical problem of plane waves in water of constant depth: *To find all harmonic functions ϕ in the infinite strip S , $-\infty < x < +\infty$, $0 \leq y \leq 1$ satisfying the boundary conditions*

$$(2.1) \quad \frac{\partial \phi}{\partial y} = py, \quad \text{for } y = 1$$

and

$$(2.2) \quad \frac{\partial \phi}{\partial y} = 0, \quad \text{for } y = 0.$$

Airy's work contains only a particular solution of this problem which is periodic in x and which has been reproduced in all textbooks.

In order to solve this problem let us consider first the eigenvalue problem given by the ordinary differential equation

$$(2.3) \quad Y'' + \lambda Y = 0, \quad Y = Y(y)$$

with the boundary conditions

$$(2.4) \quad Y' = pY, \quad \text{for } y = 1,$$

$$(2.5) \quad Y' = 0, \quad \text{for } y = 0.$$

A complete set of eigenfunctions and of corresponding eigenvalues is given by the formulas

$$(2.6) \quad Y_0 = \cosh a_0 y, \quad \lambda_0 = -a_0^2,$$

$$(2.7) \quad Y_n = \cos a_n y, \quad \lambda_n = a_n^2, \quad (n = 1, 2, \dots)$$

where a_0 is the unique (positive) root of the equation

$$(2.8) \quad a_0 \tanh a_0 = p$$

and a_1, a_2, \dots, a_n , denote the (positive) roots of the equation

$$(2.9) \quad a_n \tan a_n = p, \quad (n = 1, 2, \dots).$$

Turning back to our boundary value problem (1.1), (2.1), (2.2) we develop $\phi(x, y)$ for a fixed value of x , into the series

$$(2.10) \quad \phi(x, y) = \sum_{n=0}^{\infty} c_n(x) Y_n(y).$$

This development is possible as ϕ satisfies, for any fixed value of x , the same boundary conditions as Y_n . The Fourier coefficients c_n are given by the formulas

$$(2.11) \quad c_n(x) = C_n \int_0^1 \phi(x, y) Y_n(y) dy$$

where

$$(2.12) \quad C_n = \left(\int_0^1 Y_n^2 dy \right)^{-1}.$$

The constant C_n is the normalization factor.

From (2.11) and (1.1) it follows by differentiation that

$$c_n''(x) = -C_n \int_0^1 \phi_{yy} Y_n dy.$$

Integrating twice by parts we obtain the formula

$$c_n''(x) = -C_n [\phi_y Y_n - \phi Y_n']_0^1 - C_n \int_0^1 \phi Y_n'' dy.$$

The square bracket vanishes in view of the boundary conditions (2.1), (2.2), (2.4) and (2.5). By (2.3) we have therefore for $c_n(x)$ the differential equation

$$(2.13) \quad c_n''(x) - \lambda_n c_n(x) = 0$$

which has the following solutions:

$$(2.14) \quad c_0(x) = a_0 \cos a_0 x + b_0 \sin a_0 x$$

$$(2.15) \quad c_n(x) = a_n e^{a_n x} + b_n e^{-a_n x}, \quad (n = 1, 2, \dots).$$

On the other hand $c_n(x)$ is given by the formula (2.11).

Let us assume now that $\phi(x, y)$ satisfies the inequality

$$(2.16) \quad \int_0^1 \phi^2(x, y) dy < e^{2A|x|}, \quad A > 0$$

for $|x| \rightarrow \infty$. An application of Schwarz' inequality to the formulas

$$(2.17) \quad a_0 \cos a_0 x + b_0 \sin a_0 x = C_0 \int_0^1 \phi(x, y) Y_0(y) dy$$

$$(2.18) \quad a_n e^{a_n x} + b_n e^{-a_n x} = C_n \int_0^1 \phi(x, y) Y_n(y) dy, \quad (n = 1, 2, \dots)$$

shows immediately that $a_n = b_n = 0$ for all values of n for which a_n is greater than A , $n = 0, 1, 2, \dots$. We have therefore the following result. All solutions of our boundary value problem satisfying the inequality (2.16) are given by the formulas

$$(2.19) \quad \phi(x, y) = (a_0 \cos a_0 x + b_0 \sin a_0 x) \cosh a_0 y \\ + \sum_{n=1}^k (a_n e^{a_n x} + b_n e^{-a_n x}) \cos a_n y$$

where a and b are arbitrary constants. The exponents a_n satisfy the inequality

$$(2.20) \quad 0 < a_1 < a_2 < \dots < a_k < A < a_{k+1} < \dots$$

In particular the only bounded solution of the problem is given by the formula

$$(2.21) \quad \phi(x, y) = (a_0 \cos a_0 x + b \sin a_0 x) \cos a_0 y.$$

This solution is periodic in x and coincides with the particular solution given by Airy. The problem considered in this paragraph cannot be solved by the reduction method which will be discussed in the following section.

3. The reduction method. In this method the unknown function ϕ is replaced by a new unknown $\Phi(x, y)$ satisfying the same differential equation as ϕ but vanishing on the boundary of D . The mixed boundary problem is reduced to a problem with the classical boundary condition $\Phi = 0$.

T. Boggio [1] was the first to determine all harmonic functions ϕ satisfying the condition (1.3) on the boundary of a circle of radius one. Let us put

$$(3.1) \quad f(z) = \phi + i\psi, \quad z = x + iy = re^{i\theta}$$

where ψ denotes the conjugate function to ϕ . Since the real part of $z \frac{df}{dz}$ equals $r \frac{\partial \phi}{\partial r}$ and since the external normal to the circle has the direction of the radius r , the harmonic function

$$(3.2) \quad \Phi = r \frac{\partial \phi}{\partial r} - p\phi$$

vanishes by (1.3) on the boundary $r = 1$. Assuming that ϕ is regular for

$0 \leq r \leq 1$, we see that ϕ vanishes identically and that f satisfies therefore the ordinary linear differential equation

$$(3.3) \quad z \frac{df}{dz} - pf = ia$$

where a is a real constant. The integration of this equation shows that a regular solution ϕ exists only for $p = 1, 2, 3, \dots$, in which case ϕ is given by the formula

$$\phi = r^p (a \cos p\theta + \beta \sin p\theta).$$

A more complete analysis of the problem could have been made by the eigenvalue method, which yields also all solutions ϕ with an isolated singularity at the origin $r = 0$. (See Sec. 4.)

Recently some other interesting mixed boundary value problems corresponding to waves on sloping beaches have been discussed by Miche [8], H. Lewy [9], and J. J. Stoker [10], and others. The method of Lewy and Stoker introduces a different reduction procedure. In the following we shall discuss as an example one of the problems treated by Stoker and show that a combination of the reduction method and of the eigenvalue method yields a complete solution of the problem.

4. A mixed boundary value problem in three-dimensional wave motion.

Let us consider (see Stoker, loc. cit. [10] paragraph 9) the problem of waves in an ocean of infinite depth bounded on one side by a vertical cliff when the wave crests are not assumed to be parallel to the shore line. The corresponding boundary value problem is the following:

To find all solutions $\phi(x, y)$ of the differential equation

$$(4.1) \quad \frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} - k^2 \phi = 0$$

in the domain D , $x \geq 0$, $y \leq 0$, satisfying the boundary conditions

$$(4.2) \quad \frac{\partial \phi}{\partial y} = \phi, \quad \text{for } y = 0, \quad x > 0,$$

$$(4.3) \quad \frac{\partial \phi}{\partial x} = 0, \quad \text{for } x = 0, \quad y < 0.$$

Here k^2 denotes an arbitrary positive constant. According to Stoker we reduce the boundary conditions (4.2), (4.3) to the boundary condition $\phi = 0$ by the introduction of the function

$$(4.4) \quad \frac{\partial}{\partial x} \left(\frac{\partial}{\partial y} - 1 \right) \phi = \Phi(x, y)$$

which obviously satisfies the differential equation (4.1). It may be written in polar coordinates as follows:

$$(4.5) \quad \frac{\partial^2 \Phi}{\partial r^2} + \frac{1}{r} \frac{\partial \Phi}{\partial r} + \frac{1}{r^2} \frac{\partial^2 \Phi}{\partial \theta^2} - k^2 \Phi = 0.$$

The boundary condition is

$$(4.6) \quad \Phi = 0, \text{ for } x = 0, y < 0 \text{ and } y = 0, x > 0.$$

Instead of prescribing specifically the singularities of Φ (see Stoker, *loc. cit.*, n. 28, p. 39) we shall determine Φ by the eigenvalue method. A subsequent integration of the differential equation (4.4) will give us then all possible solutions of ϕ .

As in Sec. 2, we consider first the eigenvalue differential problem given by the equation

$$(4.7) \quad \frac{d^2 \Theta}{d\theta^2} + \lambda \Theta = 0$$

with the boundary conditions

$$(4.8) \quad \Theta = 0, \text{ for } \theta = 0 \text{ and } \theta = -\frac{1}{2}\pi.$$

A complete set of eigenfunctions and of the corresponding eigenvalues is given by the formulas

$$(4.9) \quad \Theta_n(\theta) = \sin 2n\theta, \lambda = 4n^2, \quad (n = 1, 2, \dots).$$

For a fixed value of r we have for the unknown function Φ the expansion

$$(4.10) \quad \Phi = \sum_{n=1}^{\infty} c_n(r) \sin 2n\theta$$

where

$$(4.11) \quad c_n(r) = C_n \int_{-\frac{1}{2}\pi}^0 \Phi(r, \theta) \sin 2n\theta d\theta, \quad (n = 1, 2, \dots)$$

where C_n is the normalizing factor. From this formula we find by differentiation with respect to r and by use of (4.5)

$$c_n''(r) + \frac{1}{r} c_n'(r) - \left(k^2 + \frac{4n^2}{r^2}\right) c_n(r) = \frac{C_n}{r^2} \int_0^{-\frac{1}{2}\pi} \left(\frac{\partial^2 \Phi}{\partial \theta^2} + 4n^2 \Phi\right) \sin 2n\theta d\theta.$$

The right-hand side in this equation is equal to zero as can easily be seen by two successive integrations by parts and by the use of the boundary condition (4.6). We have therefore the following differential equation for $c_n(r)$

$$(4.12) \quad c_n'' + \frac{1}{r} c_n' - \left(k^2 + \frac{4n^2}{r^2}\right) c_n = 0.$$

The general solution of (4.10) is given in terms of Bessel functions by the formula

$$(4.13) \quad c_n(r) = A_{2n} I_{2n}(kr) + B_{2n} i^{2n+1} H_{2n}^{(1)}(ikr)$$

with arbitrary real constants A_{2n} and B_{2n} . The function I_{2n} vanishes for $r = 0$ like r^{2n} but tends to infinity like $e^r r^{-1}$ for $r = \infty$. The functions $i^{2n+1} H_{2n}^{(1)}$ behave like r^{-2n} for r tending to zero and tend to zero like $e^{-r} r^{-1}$ at infinity. By the same procedure as in Sec. 2 we obtain the following results. The solutions Φ of (4.5) and (4.6) given by (4.10), can be classified according to the behaviour of the integral

$$(4.14) \quad \int_{-1}^0 \Phi^2(r, \theta) d\theta.$$

The coefficients A_{2n} in (4.13) are all equal to zero for any solution Φ for which the integral (4.14) is $o(e^{2r} r^{-1})$ at infinity. The coefficients B_{2n} vanish for $n > h$ for all solutions Φ for which the integral (4.14) is $o(r^{-2h})$ at the origin. The only solution Φ which is regular everywhere is $\Phi = 0$.

By taking $\Phi = 0$ and $\Phi = iH_2^{(1)}(ikr) \sin 2\theta$ and by integrating the corresponding differential equations (4.4) for ϕ , Stoker obtains two standing waves which can be combined into a travelling wave. One of these standing waves has a logarithmic singularity at the origin. From the results of the present paper we see the presence of a singularity is an unavoidable consequence of the linearized theory of surface waves. The contradiction of the original assumption of small amplitudes is somewhat mitigated by taking the solution with the weakest singularity at the origin. From the mathematical viewpoint, however, there is no reason to introduce any limitations on the behaviour of the solutions.

5. The method of singular integral equations. We conclude with a few remarks about this method which has been applied to the case when the domain D is a parallel strip, as in Sec. 2. Let us replace in Airy's problem the condition (2.2) by the condition

$$(5.1) \quad \phi = 0, \quad \text{for } y = 0.$$

Under certain restrictive assumptions on the behaviour of ϕ at infinity the modified problem can be reduced to a Picard integral equation [2]. However, as has been mentioned in Sec. 2, the eigenvalue method gives the solution of the same problem under less restrictive conditions. The situation is, however, different in the dock problem in a channel of finite depth, which is obtained by imposing the condition (2.1) for $y = 1$, $x > 0$ and the condition (2.2) on the remaining part of the boundary of the strip. This problem, which seems at present inaccessible by any other method, has been solved by A. E. Heins [11] by a reduction of the problem to a Wiener-Hopf equation. The assumptions which are required in order that this problem be formulated as a Wiener-Hopf integral equation are discussed in paragraph 9 of the paper by Heins. Uniqueness is studied in relation to the Wiener-Hopf integral equation to be solved. This integral equation is equivalent to the original boundary value problem subject to the conditions mentioned above. The general uniqueness theorem under less restrictive conditions, has not been discussed yet for the dock problem.

REFERENCES

- [1] T. Boggio, *Rendiconti della R. Accademia di Torino*, vol. 47 (1912), 22; see also E. Goursat, *Cours d'Analyse Mathématique*, vol. 3 (Paris, 1927), 240.
- [2] A. Weinstein, *Rendiconti della R. Accademia dei Lincei*, vol. 5, Series 6^a (1927), 259.
- [3] A. Weinstein, *C. R. Acad. Sci.*, Paris, vol. 184 (1927), 497.
- [4] G. Hoheisel, *Jber. Deutschen Math. Verein.*, vol. 39 (1930), 54.
- [5] S. Bochner, *Fouriersche Integrale* (Berlin, 1932), chap. VII.
- [6] J. L. B. Cooper, *J. London Math. Soc.*, vol. 14 (1939), 124.
- [7] A. E. Heins, *Bull. Amer. Math. Soc.*, vol. 49 (1943), 130.
- [8] A. Miche, *Annales des ponts et chaussées*, vol. 114 (1944).
- [9] H. Lewy, *Bull. Amer. Math. Soc.*, vol. 52 (1946), 737.
- [10] J. J. Stoker, *Quarterly of Applied Math.*, vol. 5 (1947), 1.
- [11] A. E. Heins, *Amer. J. Math.*, vol. 70 (1948) 730.
- [12] K. O. Friedrichs, *Communications on Appl. Math.*, vol. 1 (1948), 109.
- [13] K. O. Friedrichs and H. Lewy, *Communications on Appl. Math.*, vol. 1 (1948), 135.
- [14] F. John, *Communications on Appl. Math.*, vol. 1 (1948), 149.
- [15] E. Isaacson, *Communications on Appl. Math.*, vol. 1 (1948), 201.
- [16] G. Kreisel, *Quarterly of Applied Math.*, vol. 7 (1949), 21.

*Naval Ordnance Laboratory and
The University of Maryland*

ANGULAR MEASURE AND INTEGRAL CURVATURE

HERBERT BUSEMANN

THE Gauss-Bonnet Theorem leads through well known arguments to the fact that the integral curvature¹ of a two-dimensional closed orientable manifold M of genus p equals $4\pi(1 - p)$. This implies, for instance, that the Gauss curvature¹ K can neither be everywhere positive nor everywhere negative, if M is homeomorphic to a torus.

The relations between the sign of K and the topological structure of M have been the subject of many investigations. Those of Cohn-Vossen [4, 5] are particularly interesting, because they are not restricted to closed manifolds.

Hadamard [6] showed that the condition $K < 0$ determines to a great extent the shape of the geodesics (on closed or open manifolds). The already mentioned papers of Cohn-Vossen show also how the condition $K > 0$ influences the behaviour of the geodesics.

All these investigations rest on the Gauss-Bonnet Theorem, which states in its most primitive form that the integral curvature of a geodesic triangle equals the spherical excess of the triangle. *Thus they depend ultimately on the concept of angular measure.* This concept is in turn derived from the local, that is the Euclidean geometry, where it means amount of rotation.

The Minkowskian geometry is the local geometry of non-Riemannian metric spaces. It does not permit general rotations. If the distance is symmetric, which will always be assumed here, the Minkowskian geometry permits reflection in a point, which in the Euclidean case is equivalent to rotation through π . Therefore no particular angular measure can be entirely natural in Minkowskian geometry. This is evidenced by the innumerable attempts to define such a measure, none of which found general acceptance.

Of course, it is generally agreed that *angular measure must be additive for angles with the same vertex.* In view of our previous observation, it is natural to add the requirement that *straight angles have measure π .* It will be shown here that *any angular measure with these two properties permits us to establish for general spaces most of the above quoted results on Riemann spaces,* provided we interpret conditions like $K > 0$ on M to mean that every non-degenerate small geodesic triangle on M has positive spherical excess. For some results it is necessary to add a condition, which is always satisfied by the ordinary angles

Received October 8, 1948.

¹The English expression "total curvature" corresponds to the German "Gauss'sche Krümmung," whereas the German expression "Totalkrümmung" is "integral curvature" in English. In order not to confuse the reader interested in the original literature, the present paper avoids "total" altogether by using "Gauss curvature" corresponding to the German, and "integral curvature" corresponding to the English custom.

in Riemann spaces, and which states essentially that, in a uniform way, an angle cannot be nearly straight without having a measure close to π .

The main point of the present paper is the tenet that *angular measure in Finsler spaces is—contrary to the prevailing views—a very fruitful concept, and that it becomes unnatural and barren only through insistence on particular measures.*

The extent of the material in the Riemannian case precluded its full discussion here. Except for a glance at the connection of excess with the theory of parallels (Sec. 2) and the topological structure of compact manifolds (Sec. 3) the paper concentrates on the work of Cohn-Vossen, whose arguments are partly reproduced here.

Hadamard's results are only briefly touched because the author showed recently in [3], although not in connection with angular measure, that they do not depend on the Riemannian character of the metric.

1. Angles in systems of plane curves. In the Euclidean plane E with the (Euclidean) distance xy let S be a system of curves with the following properties:

- I. Each curve is an open Jordan curve, that is, it has a representation $q(t)$, $-\infty < t < \infty$ where $q(t)$ is continuous and $q(t_1) \neq q(t_2)$ when $t_1 \neq t_2$.
- II. $q(t)q(0) \rightarrow \infty$ when $|t| \rightarrow \infty$.
- III. Any two distinct points of E lie on exactly one curve of S .

The curve in S (S -curve) determined by the two points a, b will be denoted by $g(a, b)$. On $g(a, b)$ the points a, b bound an arc $t(a, b)$. The symbol (acb) means that a, b, c are three different points and that c lies on $t(a, b)$. We also put $t(a, a) = a$.

The S -curves satisfy all the axioms of order and connection of Hilbert, in particular the axiom of Pasch.² In addition, $a_r \rightarrow a$ and $b_r \rightarrow b$ implies $t(a_r, b_r) \rightarrow t(a, b)$ and, if $a \neq b$, also $g(a_r, b_r) \rightarrow g(a, b)$. The arrow indicates here that $t(a, b)$ or $g(a, b)$ is Hausdorff's closed limit of the sets $t(a_r, b_r)$ or $g(a_r, b_r)$.² A point p of an S -curve g divides g into two (closed) rays r_1, r_2 , which we call opposite.

If r_1 and r_2 are two different rays with the same origin p , then $r_1 \cup r_2$ divides E into two (closed) domains D_1 and D_2 . The sets of all rays with origin p in D_1 and D_2 respectively are the two angles with legs r_1 and r_2 . They are called straight if r_1 and r_2 are opposite. Otherwise exactly one of the domains is S -convex³ and we call the corresponding angle the convex angle $(r_1, r_2)^{\text{vex}}$, and the other the concave angle $(r_1, r_2)^{\text{cav}}$. It is convenient to complete this definition by letting $(r_1, r_1)^{\text{vex}}$ mean the set consisting of r_1 alone and $(r_1, r_1)^{\text{cav}}$ the set of all rays with origin p . If a, b, c are three points not on one S -curve, then $\angle abc$ means the convex angle whose legs are the rays from b through a and c .

²Proofs are found in [1, Sec. III.3].

³A set X is S -convex if $a, b \in X$ implies $t(a, b) \subset X$. Compare [1, Sec. III.3].

We now assume that an *angular measure* $|D|$ has been defined for the angles D in S with the following properties:

- 1) $|D| \geq 0$.
- 2) $|D| = \pi$ if and only if D is straight.
- 3) If D_1 and D_2 are two angles with a common leg but with no other common ray, then $|D_1 \cup D_2| = |D_1| + |D_2|$.

We say that the angle D_r tends to the angle D , if the legs of D_r tend to the legs of D , and if $r, \epsilon D_r$ and $r, \rightarrow r$ implies $r, \epsilon D$. We call the angular metric *continuous* if

- 4) $D_r \rightarrow D$ implies $|D_r| \rightarrow |D|$.

Some consequences of 1), 2), 3) are

- a) $|D| = 0$ if and only if the legs of D coincide.

For if the legs r_1, r_2 of D coincide, let D' denote one of the two straight angles with $r_1 = r_2$ as one leg. Then $D \cup D' = D'$, therefore by 2) and 3)

$$\pi = |D \cup D'| = |D| + |D'| = |D| + \pi,$$

so that $|D| = 0$. Conversely let $|D| = 0$. Its legs r_1 and r_2 cannot be opposite by 2). Denote the opposite ray to r_1 by r_3 . If r_1 and r_2 did not coincide and $D = (r_1, r_2)^{\text{conv}}$ then by 1), 2), 3) $|D| = \pi + |(r_3, r_2)^{\text{conv}}| > \pi$. If $D = (r_1, r_2)^{\text{conv}}$ then $\pi = |D| + |(r_3, r_2)^{\text{conv}}| = |(r_3, r_2)^{\text{conv}}|$ although $(r_3, r_2)^{\text{conv}}$ is not straight.

- b) Convex angles have measure less than π and conversely.

Concave angles have measure greater than π and conversely.

- c) Vertical angles are equal.

- d) The sum of the measures of the angles in a triangle abc (set bounded by three segments $t(a, b)$, $t(b, c)$, $t(c, a)$, where a, b, c are not on one S -curve) is positive and less than 3π .

- e) If the angular metric is continuous and the points a, b, c , are not on an S -curve and tend to a point p , then the sum of the angular measures in the triangle a, b, c , tends to π .

A proof follows immediately from the observation that $g(a, b_r)$, $g(b_r, c_r)$ and $g(c_r, a_r)$ may be assumed to converge. Then $\angle b_r a_r c_r$ and the vertical angles to $\angle a_r b_r c_r$ and $\angle a_r c_r b_r$ tend to three angles whose union is a straight angle.

Some of the preceding remarks extend in the usual way to degenerate triangles and will be used for such triangles.

The excess $\epsilon(abc)$ of the triangle abc is defined as

$$(1) \quad \epsilon(abc) = |abc| + |bca| + |cab| - \pi,$$

where $|abc| = |\angle abc|$.

Degenerate triangles have excess 0. If the triangle abc is decomposed (simplicially by S -curves) into the triangles a, b, c_r , then

$$\epsilon(abc) = \Sigma \epsilon(a, b, c_r).$$

If a_1, \dots, a_n are the vertices of a simple closed polygon P with sides $t(a_i, a_{i+1})$ and α_i is the measure of the angle at a_i measured inside the closed⁴

⁴In this paper domains bounded by geodesic polygons are always understood to be closed.

domain G bounded by P then for any simplicial subdivision of G (by S -curves is always understood) into triangles a, b, c ,

$$(2) \quad \epsilon(a, b, c) = 2\pi - \Sigma(\pi - \alpha_i) = \Sigma \alpha_i - (n - 2)\pi.$$

Let r_1 and r_2 be two opposite rays with origin p determining the two straight angles D_1 and D_2 . If $a_i \in r_i$, $a_i \neq p$ and $q \in D_1 - (r_1 \cup r_2)$ then a ray r_x with origin p and through a point $x \in t(a_1, q) \cup t(q, a_2)$ traverses monotonically all rays in D_1 as x traverses $t(a_1, q) + t(q, a_2)$. Therefore $|(r_1, r_x)^{\text{vex}}| = \phi(r_x)$ is a strictly increasing function with $\phi(r_1) = 0$, $\phi(r_2) = \pi$. The values of $\phi(r_x) = |(r_1, r_x)^{\text{cav}}|$ for $r_x \in D_2$ are determined by 2). If D is any angle with vertex p which does not contain r_1 and has legs r', r'' , then

$$(3) \quad |D| = |\phi(r') - \phi(r'')|.$$

If D contains r_1 , then

$$(3') \quad |D| = 2\pi - |\phi(r') - \phi(r'')|.$$

Conversely, if in D_1 any strictly increasing function $\phi(r_x)$ with $\phi(r_1) = 0$, $\phi(r_2) = \pi$ is given, and $\phi(r_x)$ is determined in D_2 to satisfy 2), then (3) and (3') determine an angular measure at p which satisfies 1), 2), 3).

2. Excess and parallels. The present section is concerned with the relation of the angular metric in a system S to the theory of parallels. It will not be needed later on but will elucidate the meaning of an angular metric.

If g^+ is an oriented S -curve and x traverses g^+ in the positive sense, then the line $g(p, x)$ converges for any fixed point p to a line a . If $g^+(a, b)$ denotes generally the line $g(a, b)$ with the orientation in which b follows a then $g^+(p, x)$ tends to an orientation a^+ of a . We call $a(a^+)$ the (oriented) asymptote to g^+ through p (for a proof of this and the next statements see [1, Sec. III.3]). The line a does not intersect g . The asymptote to g^+ through any point $q \in a$ is again a . But in general g^+ is not an asymptote to a^+ , for an example see [1, Sec. III.5].

Let the *parallel axiom* hold, that means, through a given point p not on a given line g there is exactly one line h which does not intersect g . If we determine angular measure at one point p as at the end of the preceding section but with a continuous ϕ , and define measure for an arbitrary angle as equal to the corresponding angle at p with legs parallel to the given angle, then condition 4) is also satisfied and the excess of any triangle is 0.

(4) *A system S in which the parallel axiom holds possesses continuous angular metrics with excess 0.*

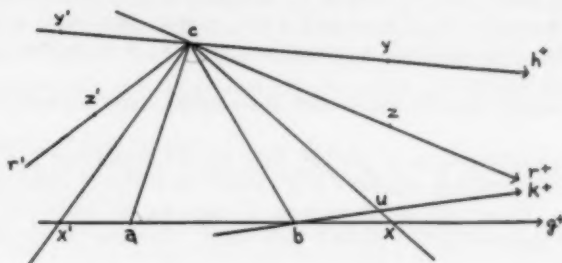
However, it is not true that zero excess implies the parallel axiom nor does the parallel axiom imply that every continuous angular metric has excess 0. The only statement which holds without further conditions on the angular metric is the following:

(5) *If the excess is non-positive and the angular metric is continuous then the parallel axiom implies zero excess.*

Let abc be a non-degenerate triangle. If (abx) and x traverses $g^+ = g^+(a, b)$

then $|bac| + |acx| < \pi$ and $g^+(c, x)$ tends to the asymptote h^+ to g^+ . If y follows c on h^+ then $|acx| \rightarrow |acy| = |acb| + |bcy|$ so that

$$(6) \quad |bac| + |acy| \leq \pi.$$



If (ycy') and (bax') then $g^+(c, y')$ is, because of the parallel axiom, the asymptote to $g^+(b, a)$ through c . As before we see

$$(6') \quad |x'ac| + |acy'| \leq \pi$$

and since $|x'ac| = \pi - |bac|$, $|acy'| = \pi - |acy|$ it follows from (6) and (6') that $|bac| + |acy| = |x'ac| + |acy'| = \pi$ so that $|bac| = |acy'|$. For the same reason $|abc| = |bcy|$ and since $|acy'| + |acb| + |bcy| = \pi$ the theorem is proved.

It is clear that this argument and the additivity of the excess yield the following more general fact:

(7) *If the excess is non-positive and the metric is continuous and there is only one line h through a point p not on g which does not intersect g , then any triangle, whose vertices are in the closed strip bounded by g and h , has excess 0.*

The arguments in the above proof could be reversed if $|cxb| \rightarrow 0$. The following examples will show that further progress is impossible without this property. A hemisphere H without the bounding great circle can be mapped on the Euclidean plane E in such a way that the arcs of great circles in H go into the Euclidean straight lines in E . If we assign to an angle in E the measure of the corresponding spherical angle (in the usual sense), then the excess in any triangle of E is positive in spite of the parallel axiom. With this measure the same holds for the straight line pieces in the interior of a circle in E . On the other hand the Euclidean angles in E may be used as angular measure for those same pieces. This means that both positive and zero excess are compatible with the hyperbolic parallel axiom.

We call an *angular metric* in a system S of curves *complete* if it is continuous and $|pxq| \rightarrow 0$ whenever x traverses a ray with origin p from p toward ∞ .

(8) *In a complete angular metric the excess cannot always be positive.*

With the same notation as above positive excess would yield $\epsilon(x'cx) > \epsilon(abc) > \pi$. Because of $|cax'| \rightarrow 0$ and $|cx'x| \rightarrow 0$ it would follow that for x and x' which are sufficiently far away $|x'cx| > \pi$ which is impossible.

(9) *In a complete angular metric zero excess implies the parallel axiom.*

For then $|cax| + |acx| + |axc| = \pi$ and $|axc| \rightarrow 0$. Moreover $t(c, x)$ tends to a ray r which lies on the asymptote r^+ through c to g^+ . If $z \in r$, $z \neq c$ then $|acx| \rightarrow |acz|$, hence $|cax| + |acz| = \pi$. Similarly $t(c, x')$ tends to ray r' on the asymptote through c to $g^+(b, a)$ and if $z' \in r'$, $z' \neq c$, then $|cax'| + |acz'| = \pi$. It follows from $|cax| + |cax'| = \pi$ that $|acz| + |acz'| = \pi$, so that r and r' are opposite rays, q.e.d.

(10) *In a complete angular metric with non-positive excess asymptotes are symmetric.*

If r^+ is an asymptote to g^+ and g^+ were not an asymptote to r^+ then the asymptote to r^+ through a point b of g^+ would be a line f^+ different from g^+ (see Figure). If $c \in r^+$ and u follows b on f^+ then $g(c, u)$ intersects g^+ by the definition of asymptotes in a point x with (cux) . Because the excess in bux is non-positive

$$|cub| \geq |ubx| + |bxu| > |ubx| > 0$$

but $|cub| \rightarrow 0$ when u traverses f^+ in the positive sense.

Example 1) in [1, Sec. III. 5] yields, with the ordinary Euclidean angles, a complete angular metric and non-symmetric asymptotes which shows that (10) would not hold without the assumption that the excess is non-positive.

(9) and (10) are first examples of statements which connect conditions on the excess with topological properties (in this case of the system S).

3. Angular measure for curve systems on two-dimensional manifolds. The word surface will be used here to denote a connected two-dimensional topological manifold.

As in Sec. 2 for the plane we consider on a given surface M a system S of curves with the topological properties of geodesics. The existence of such a system is guaranteed by the following two conditions.

1) *Every point p of M has a neighbourhood $U(p)$ homeomorphic to the plane, in which a system S_p of curves is distinguished with the properties I, II, III of Sec. 1.*

2) *If a, b, c lie in $U(p) \cap U(q)$ then (abc) holds with respect to S_p if and only if it holds with respect to S_q .*

By 2) a segment $t(a, b)$ in S_p is also a segment in S_q . Therefore the notation $t(a, b)$ can be used without reference to a definite system S_p as long as a and b both lie in some $U(p)$.

The concept of a geodesic will actually not be used in the sequel. But since 1) and 2) are derived from this concept, we mention that an S -geodesic is to be defined as a continuous curve $x(t)$, $-\infty < t < \infty$ with the following property: if t_0 is given and $x(t_0) \in U(p)$ then a suitable subarc $t_1 < t < t_2$ with $t_1 < t_0 < t_2$ of $x(t)$ represents a curve in S_p . The existence of geodesics can be established by the procedure of [2, Sec. II.5].

If $a \in U(p) \cap U(q)$ then a ray r_p with origin a in S_p will, in general, not be a ray in S_q , but by 2) the ray r_p either contains, or is contained in, a ray r_q with origin a of S_q , which is uniquely determined by r_p .

If r_p^1, r_p^2 are two rays with origin a in S_p and r_q^1, r_q^2 are the corresponding rays in S_q , then 2) clearly implies that r_p^1 and r_p^2 are opposite if and only if r_q^1 and r_q^2 are. Also, if r_p^1 and r_p^2 are not opposite, then a ray in $(r_p^1, r_p^2)^{\text{vex}}$ or $(r_p^1, r_p^2)^{\text{cav}}$ corresponds to a ray in $(r_q^1, r_q^2)^{\text{vex}}$ or $(r_q^1, r_q^2)^{\text{cav}}$ respectively.

These facts lead to the following *formal definition of a ray* r with origin a in S : r is a set of rays in the local curve systems with these properties:

a) r contains exactly one ray of every S_p for which $a \in U(p)$ and no ray of any other S_p .

b) If $r_p \in r$ and $r_q \in r$ then either $r_p \supset r_q$ or $r_p \subset r_q$.

The meaning of angles in S , of symbols like $(r_1, r_2)^{\text{vex}}$, and of convergence of angles is now obvious.

An *angular measure for the angles in S* is then characterized by the properties 1), 2), 3) of Sec. 1 and a continuous angular measure by the additional property 4). Through the natural one-to-one correspondence between the angles with vertex a in S and the angles with vertex a in S_p , $U(p) \ni a$, an angular measure in S induces an angular measure in S_p .

Whenever the word triangle is used it is understood that its vertices lie in one $U(p)$. The excess of a triangle is still defined by (1). A geodesic polygon

P on M is a curve of the form $\bigcup_{i=1}^n t(a_i, a_{i+1})$, $a_i \neq a_{i+1}$. Some of the angles of P may be straight, that is the segments $t(a_{i-1}, a_i)$ and $t(a_i, a_{i+1})$ may belong to opposite rays with origin a_i .

We then call a_i an *improper vertex* of P , otherwise a *proper vertex*. If all vertices of P are improper then P is a *geodesic arc*. If in addition $a_1 = a_{n+1}$ and the angle at a_1 is straight, we call P a *closed geodesic*.

Let G be a compact domain of finite genus on M which is bounded by n simple closed mutually non-intersecting geodesic polygons. If G is simplicially divided into triangles, then the number of vertices minus the number of sides plus the number of triangles is an integer $X(G)$ which depends only on G and not on the choice of simplicial division. According to the terminology prevailing in topology, $X(G)$ is the *negative Euler characteristic* of G .⁵

Any simplicial division of G into triangles a, b, c , satisfies the following fundamental relation

$$(11) \quad \sum_p \epsilon(a, b, c_p) = 2\pi X(G) - \sum_i (\pi - \alpha_i),$$

where α_i are the angular measures of the angles at the vertices of the boundary of G measured in G .⁶ It is immaterial whether α_i traverses the angles at all or only the proper vertices.

If M is compact and has finite genus p we find that for any simplicial decomposition of M into triangles a, b, c ,

⁵Compare Kerekjarto [8] and Seifert-Threlfall [9]. Cohn-Vossen [4] calls $X(G)$ (and not $-X(G)$) the characteristic of G .

⁶A modification of the topological proof for (11) which is adapted to the present conditions is found in [4, p. 120].

$$(12) \quad \sum \epsilon(a, b, c_r) = 2\pi X(M) = \begin{cases} 4\pi(1 - p) & \text{if } M \text{ is orientable,} \\ 2\pi(2 - p) & \text{if } M \text{ is not orientable.} \end{cases}$$

The number $\sum \epsilon(a, b, c_r)$ in (11) or (12) which is independent of the simplicial division is called the integral curvature $C(G)$ of G or $C(M)$ of M . We say that M or a domain G on M has positive, negative, non-positive, non-negative, or zero curvature if for every non-degenerate triangle abc in M or G

$$\epsilon(abc) > 0, < 0, \leq 0, \geq 0, \text{ or } = 0 \text{ respectively.}$$

If the curvature of a two dimensional Riemann space R is non-negative, non-positive, zero, positive, or negative in the usual sense then R has the same property in the present sense. The converse is true in the first three cases, but not always in the last two. If the Gauss curvature of R is positive (negative) except on some curves or isolated points, R has still positive (negative) curvature in the present sense. The existence part of the following theorem follows therefore from well-known facts regarding Riemann spaces, the remainder is a consequence of (12).

(13) THEOREM. A compact surface M can be provided with a system S of geodesics and an angular measure such that curvature is:

- non-negative, if and only if M is homeomorphic to the sphere, torus, one-sided torus (also called Klein-Bottle), or the projective plane.
- non-positive, if and only if M is not homeomorphic to the sphere or the projective plane.
- positive, if and only if M is homeomorphic to the sphere or the projective plane.
- negative, if and only if M is not homeomorphic to the sphere, torus, one-sided torus or the projective plane.

A torus or one-sided torus with non-positive or non-negative curvature has curvature 0.

4. Two dimensional metric manifolds. No statements which approach (13) in completeness seem to be possible for non-compact surfaces unless the curves in S are really geodesics in the metric sense, and not only curves with the topological properties of geodesics. That M is a space with metric geodesics is expressed by the following conditions:

- M is a metric space with distance xy .
- M is finitely compact, or a bounded sequence has an accumulation point. The fact that the three points a, b, c are different and satisfy the relation $ab + bc = ac$ will be written as (abc) .
- M is convex, that is, for any two distinct points a, c a point b with (abc) exists.
- Prolongation is locally possible, or for every point p there is a $\rho(p) > 0$ such that for any two different points a_1, a_2 with $a, p < \rho(p)$ a point d with $(a_1 a_2 d)$ exists.
- Prolongation is unique, or, if $(a_1 a_2 d')$, $(a_1 a_2 d'')$ and $a_2 d' = a_2 d''$ then $d' = d''$.

These axioms guarantee the existence of geodesics (compare [2]). In the present case we add

F. M has dimension 2 (in the sense of Menger-Urysohn).

It can be proved that M is a connected topological manifold or a surface (for this and the following statements see [1, Sec. I.4]). A space which satisfies Axioms A to F will be called a G -surface.

A *metric segment* is an isometric map of a Euclidean segment. If $U(p)$ is the interior of a sufficiently small geodesic triangle on M , then the open metric segments in $U(p)$ with endpoints on the boundary of $U(p)$ form a curve system S_p with properties 1) and 2) of Sec. 3.

Since any two points of M can be connected by a metric segment, only those metric segments are segments in the previous sense, which lie entirely in one $U(p)$. But since every metric segment can be divided into a finite number of metric segments each of which lies in one $U(p)$, and the angles at the points of division are straight, the distinction between the two kinds of segments turns out to be immaterial and will therefore be dropped.

If M has finite connectivity it can be represented topologically as a compact manifold \bar{M} of finite genus which has been punctured at a finite number of points z_1, \dots, z_k . Let P be a simple closed geodesic polygon which bounds on \bar{M} a simply connected closed domain T which contains exactly one z_i , say z_{i_0} . Because of B the set $T - z_{i_0}$ appears on M as a set which looks like a half cylinder and extends to ∞ . We call T a *tube* (Fluchtgebiet in the terminology of Cohn-Vossen [4]).

The tubes are the new feature of non-compact M as compared to compact surfaces. The study of non-compact M must therefore be based on the properties of tubes. *The remainder of this section investigates tubes.*

With the above notation, consider on T the class $C(u)$, $u > 0$, of all curves C which are homotopic to P on T and have distance at most u from P . Whether this distance is measured on M or on T is immaterial. For if measured on M then a segment connecting a point of P to a point of C exists whose length equals the distance of P and C on M . This segment cannot contain a second point of P and lies therefore entirely in T .

$C(u)$ contains curves of finite length (for instance P). Since T , considered as space, satisfies B and every member of $C(u)$ contains a point whose distance from P is at most u , there is a shortest curve $R(u)$ in $C(u)$ (for a proof compare [1, p. 10] and [2, p. 234]). The length $\lambda(u)$ of $R(u)$ is obviously a non-increasing function of u and the triangle inequality yields easily that $\lambda(u)$ is continuous (see [4, §16]). We represent $R(u)$ with the arc length t as parameter in the form $x(t)$, $0 \leq t \leq \lambda(u)$, $x(0) = x(\lambda(u))$. Notice first

(14) If $x(t_0)$ is not a vertex of P and has either distance greater than u from P or is not the only point of $R(u)$ whose distance from P is at most u then the subarc⁷ $t_0 - \delta \leq t \leq t_0 + \delta$ of $x(t)$ is a segment for sufficiently small $\delta > 0$.

⁷This inequality is to be replaced by the two inequalities $0 \leq t \leq \delta$ and $\lambda(u) - \delta \leq t \leq \lambda(u)$ if $t_0 = 0$ or $t_0 = \lambda(u)$.

For otherwise the subarc $t_0 - \delta \leq t \leq t_0 + \delta$ can be replaced by a segment with the same endpoints. If $\delta > 0$ is small enough, the new curve R' will still lie in T , even when $x(t_0)$ lies on P , but is not a vertex of P . Moreover R' will still be homotopic to P and have distance $< u$ from P . But the length of R' would be less than $\lambda(u)$ which contradicts the definition of $R(u)$.

(14) implies that $R(u)$ is a geodesic polygon. Moreover, if $R(u)$ contains points with distance $< u$ from P then $R(u)$ contains infinitely many such points and none of them can be a vertex of $R(u)$. Therefore we see

(15) $R(u)$ is either a closed geodesic, or all its vertices are vertices of P , or $R(u)$ has exactly one vertex and its distance from P equals u , whereas all other points of $R(u)$ have greater distance from P than u .

We show next that $R(u)$ is a Jordan curve. Since $R(u)$ is homotopic to P and T is homeomorphic to a halfcylinder, $R(u)$ must contain subpolygon R' which is a Jordan curve and homotopic to P . If $R \neq R'$ then R' cannot have distance $\leq u$ from P , otherwise R' would belong to $C(u)$ and be shorter than $R(u)$. Therefore $R(u) - R'$ contains a point r with distance $\leq u$ from P ; r may be chosen as $x(0)$. Then R' is a subarc of $x(t)$ of the form $0 < \alpha \leq t \leq \beta < \lambda(u)$ with $x(\alpha) = x(\beta)$. The arcs $0 \leq t \leq \alpha$ and $\beta \leq t \leq \lambda(u)$ of $x(t)$ must have the same length, otherwise replacing the longer by the shorter would yield a curve in $C(u)$ with smaller length than $\lambda(u)$.

Replacing the arc $\beta \leq t \leq \lambda(u)$ by the arc $0 \leq t \leq \alpha$, that is defining $y(t) = x(t)$ for $0 \leq t \leq \beta$ and $y(t + \beta) = x(\alpha - t)$ for $0 \leq t \leq \alpha = \lambda(u) - \beta$, yields again a curve R^* in $C(u)$ of length $\lambda(u)$. Statement (14) would then apply to R^* , hence for small $\delta > 0$ the arcs $\alpha - \delta \leq t \leq \alpha + \delta$ and $\beta - \delta \leq t \leq \beta + \delta$ would be segments. By construction the arcs $\alpha - \delta \leq t \leq \alpha$ and $\beta \leq t \leq \beta + \delta$ coincide. The uniqueness of the prolongation E would imply that the arcs $\alpha \leq t \leq \alpha + \delta$ and $\beta - \delta \leq t \leq \beta$ also coincide, but then R' would not be a Jordan curve.

Since $R(u)$ is a simple closed geodesic polygon homotopic to P it bounds a subtube $T(u)$ of T . If a vertex r of $R(u)$ is a vertex of P , then the angle of $R(u)$ at r measured in $T(u)$ cannot be convex, otherwise $R(u)$ could, because of $u > 0$, be shortened without violating the conditions for belonging to $C(u)$.

Finally it will be proved that in case $R(u)$ has exactly one vertex q with distance u from P , the angle at q measured in $T(u)$ must be convex. Let t be a segment of length u connecting q to a point d on P . Then t cannot contain other points of either P or $R(u)$ because the distance of $R(u)$ from P would then be smaller than u , contrary to (15). If the angle D at q in $T(u)$ were concave, let c_1, c_2 be points on the legs of D and close to q . Then the interior I of the triangle qc_1c_2 would lie outside of $T(u)$. Also, $I \cup T(u)$ contains a neighbourhood of q . The segment t connects d to q without entering $T(u)$. It must therefore cross $t(c_1, c_2)$ at a point q' and q' has distance $u' < u$ from P . If then the arc $t(c_1, q) \cup t(q, c_2)$ of $R(u)$ is replaced by $t(c_1, c_2)$, the length decreases so that $\lambda(u') < \lambda(u)$, which is impossible.

Thus we have proved the Theorem of Cohn-Vossen:

(16) $R(u)$ is a simple closed geodesic polygon. It is either a closed geodesic, or all

its proper vertices are also vertices of P and the corresponding angles measured in $T(u)$ are concave, or $R(u)$ has exactly one proper vertex q , which is the only point on $R(u)$ with distance u from P and the angle at q in $T(u)$ is convex.

5. Angular metric and structure of non-compact metric surfaces. We now assume that an angular measure has been defined for the system of geodesics of a G -surface M of finite connectivity. With the notations of the preceding section we associate with the points z_i a set of k mutually disjoint tubes T_i , each bounded by a geodesic polygon P_i .

Let $u_i > 0$. By Cohn-Vossen's Theorem T_i contains a subtube $T_i(u_i)$ bounded by a geodesic polygon $R_i(u_i)$ such that $R_i(u_i)$ is either a closed geodesic, or all angles of $R_i(u_i)$ measured in $T_i(u_i)$ are concave, or $R_i(u_i)$ has exactly one convex angle whose measure in $T_i(u_i)$ is not zero because $R_i(u_i)$ is a Jordan curve. Let $k' (\leq k)$ denote the number of the $R_i(u_i)$ with a convex angle.

Call G the compact domain on M bounded by the $R_i(u_i)$. Since concave (convex) angles of $R_i(u_i)$ measured in $T_i(u_i)$ are convex (concave) when measured in G , the relation (11) yields

$$(17) \quad C(G) \leq 2\pi X(G) + k'\pi,$$

where the equality sign holds only when all $R_i(u_i)$ are closed geodesics.

It is well-known (see [8, pp. 145, 147]) that

$$X(G) = \begin{cases} 2 - (2p + k), & p \geq 0 \text{ if } M \text{ is orientable,} \\ 2 - (p + k), & p \geq 1 \text{ if } M \text{ is non-orientable,} \end{cases}$$

where p is the genus of M or G .

Hence for non-compact M (that is $k \geq 1$) and $C(G) \geq 0$ only the following cases are possible. If M is orientable, then $p = 0$ and 1) $k = 1, k' = 0, 1$; 2) $k = 2, k' = 0, 1, 2$; 3) $k = k' = 3$. If M is not orientable then $p = 1, k = 1, k' = 0, 1$.

Taking first only k into account we find in addition to Theorem (13):

(18) THEOREM. *A non-compact G -surface with non-negative curvature is homeomorphic to a plane, a cylinder, a sphere with three holes, or a Moebius strip.*

This agrees again with the known facts regarding Riemann spaces, except for the sphere with three holes. It may therefore be of interest to discuss this exception in some detail.

For that and other purposes we divide the tubes, following Cohn-Vossen, into three categories. Let T be a tube bounded by the simple closed geodesic polygon P , β the greatest lower bound of the length of all curves homotopic to P on T . We call minimal sequence a sequence of curves on T homotopic to P whose length tends to β .

If there is no bounded minimal sequence, we call T *contracting*.

If no subtube of T is contracting we call T *expanding*.

If T is neither contracting nor expanding we call T *bulging*.^a

^aCohn-Vossen calls a contracting tube a *Schaft*, and uses *Kelch* for both bulging and expanding tubes. The latter are distinguished as "*eigentliche Kelche*."

The following facts are obvious (Compare [4, §18]):

- (19) A subtube of a contracting tube is contracting.
- (20) A subtube of an expanding tube is expanding.
- (21) A subtube of a bulging tube which is sufficiently far away is contracting.

An expanding or bulging tube contains a bounded minimal sequence. This sequence contains a converging subsequence which tends to a curve R homotopic to P of length β (see [1, Sec. I.1]). If the distance of R from P is u' then $R(u) = R$ for every $u > u'$. By the Theorem of Cohn-Vossen $R(u)$ is either a closed geodesic or all its angles measured in $T(u)$ are concave.

In the preceding discussion k' may therefore be interpreted as the number of contracting tubes and we see:

- (22) A sphere with three holes and non-negative curvature has only contracting tubes and the angle of at least one $R_i(u_i)$ measured in $T_i(u_i)$ must be less than $\pi/3$.

Cohn-Vossen proves that u can be chosen such that the angle of an $R(u)$ on a contracting tube is as close to π as desired. This is not true for general angular metrics.

An instructive example can be obtained as follows: In the ordinary space consider the surface M of revolution $z = (x^2 + y^2)^{-1}$. It is homeomorphic to a cylinder or a sphere with two holes, one corresponding to $z = \infty$, the other to $z = 0$. If P_1 and P_2 are two simple closed geodesic polygons associated with those holes as in the beginning of this section, say P_1 to $z = \infty$ and P_2 to $z = 0$, and T_i is the tube bounded by P_i , then P_1 is contracting and P_2 is expanding. Well-known facts on geodesics on surfaces of revolution yield readily that the $R_1(u)$ have all exactly one convex angle $D(u)$ in $T_1(u)$ whose vertex $q(u)$ has distance u from P . Because M is a surface of revolution and the meridians are geodesics the $q(u)$ either lie, or can be assumed to lie, on one meridian.

Let $\alpha(u)$ be the ordinary radian measure of $D(u)$; by Cohn-Vossen's already mentioned result $\alpha(u) \rightarrow \pi$ for $u \rightarrow \infty$. We now define an angular measure at $q(u)$ as follows. If $D(u) = (r_1, r_2)^{\text{vex}}$ let D be the straight angle of the form (r_1, r'_1) that contains $D(u)$. For any $r \in D$ let $\alpha(r)$ be the ordinary radian measure of $(r_1, r)^{\text{vex}}$, so that $\alpha(r_2) = \alpha(u)$ and define $\phi(r)$ by

$$\phi(r) = \begin{cases} \delta \alpha(r) & \text{for } r \in D(u), \quad 0 < \delta < 1, \\ \delta \alpha(u) + (\alpha(r) - \alpha(u)) \cdot (\pi - \delta \alpha(u)) \cdot (\pi - \alpha(u))^{-1} & \text{for } r \in D - D(u). \end{cases}$$

We use $\phi(r)$ as at the end of Sec. 1 to define an angular measure at the point $q(u)$.

For points on the same parallel circle as $q(u)$ we define angular measure in an obvious way by rotation of $q(u)$ about the z -axis. On the remainder of M we use the ordinary angular metric. Then the new angular metric is continuous on M except on the parallel circle corresponding to $u \rightarrow 0+$. It can easily be smoothed out there.

Then $D(u) = \delta\pi$ for all $u > 0$, so that $D(u)$ does not approach π for $u \rightarrow \infty$.

By the same method a sphere with three contracting tubes can be constructed for which the angles of all $R_i(u_i)$ are less than $\pi/3$, so that (22) cannot be improved without a new condition on the angular metric. The example shows also in which direction such a condition has to go:

An angular metric is called *uniform* on a subset G of M if two positive functions $\delta(\epsilon)$ and $\rho(p, \epsilon)$, where $0 < \epsilon < 1$ and $p \in G$, exist, such that the relations $0 < a_1 p = a_2 p < \rho(p, \epsilon)$ and $a_1 a_2 / (a_1 p + p a_2) \geq 1 - \delta(\epsilon)$ imply for $p \in G$ that $|a_1 p a_2| \geq \pi - \epsilon$.

The uniformity is contained in the requirement that $\delta(\epsilon)$ is independent of p . The usual angular metric of a Riemann space is uniform, because $|a_1 p a_2| \rightarrow 2 \arccos [(1 - \delta)/2]$ for $a_i \rightarrow p$ and $a_1 a_2 / (a_1 p + p a_2) = 1 - \delta$. According to Cohn-Vossen (16) may be completed by

(23) THEOREM. *If the angular metric on the tube T is uniform, then for a suitable $u_0 > 0$ the curve $R(u_0)$ is either a closed geodesic, or all angles of $R(u_0)$ measured in $T(u_0)$ are concave, or the angle of $R(u_0)$ at its only vertex q is at least $\pi - \epsilon$.*

Proof. Consider the function

$$f(u) = \lambda(u) + 2\delta(\epsilon)u, \quad u \geq 1.$$

Since $\lambda(u)$ is non-negative and continuous, $f(u)$ reaches its minimum at some value $u_0 (\geq 1)$. Therefore

$$\begin{aligned} \lambda(u_0 + h) + 2\delta(\epsilon)(u_0 + h) &\geq \lambda(u_0) + 2\delta(\epsilon)u_0, \quad \text{for } h > 0, \\ (23a) \quad 2\delta(\epsilon)h &\geq \lambda(u_0) - \lambda(u_0 + h), \quad \text{for } h > 0. \end{aligned}$$

If $R(u_0)$ is not a closed geodesic or its angles are not concave, let q_0 be the vertex of $R(u_0)$. If $t(q_0, a^*_{i1}), t(q_0, a^*_{i2})$ are proper segments on the legs of the angle at q_0 , let $(q_0 a_i a_i^*)$, $i = 1, 2$, with $h = q_0 a_i < \rho(q_0, \epsilon)$.

Consider the curve R' originating from $R(u_0)$ by replacing $t(a_1, q_0) \cup t(q_0, a_2)$ by $t(a_1, a_2)$. The distance of R' from P is at most $u_0 + h$. Therefore $\lambda(u_0 + h) \leq \lambda' = \text{length of } R'$ and

$$(24) \quad \lambda(u_0) - \lambda(u_0 + h) \geq \lambda(u_0) - \lambda' = 2h - a_1 a_2 = 2h(1 - a_1 a_2 / 2h),$$

and (23a) and (24) yield

$$\delta(\epsilon) \geq 1 - a_1 a_2 (a_1 q_0 + q_0 a_2)^{-1};$$

hence $|a_1 q_0 a_2| \geq \pi - \epsilon$ by the definition of $\delta(\epsilon)$.

From (22) and (23) we find

(25) *A sphere with three holes and uniform angular metric cannot have non-negative curvature.*

Other well-known theorems can be proved under these general conditions. We mention only one example from Hadamard's theory (see [6]):

(26) *On a G-surface M with negative curvature a class of freely homotopic curves contains at most one closed geodesic.*

The universal covering space of a G-surface M is again a G-surface M' (see [2, Sec. 13]). An angular metric on M induces an angular metric on M' . If M has negative curvature, then M' has negative curvature with respect to this induced metric.

If M contained two freely homotopic closed geodesics g_1 and g_2 , draw a segment t from a point p_1 of g_1 to a point p_2 of g_2 . The figure consisting of g_1 , g_2 and t is image of a quadrangle in M' whose angle sum is 2π . Because of the additivity of the excess M' must contain arbitrarily small non-degenerate triangles with non-negative excess, but then M would contain such triangles.

6. The integral curvature of non-compact surfaces. A polygonal region G is the closure of an open set on M whose boundary B (if any) is locally a simple geodesic polygon. That means: if p is any point on B then a geodesic triangle abc exists which contains p in its interior I and such that the intersection of B with the closure of I decomposes I and consists of two segments $t(p, x)$, $t(p, y)$.

For compact G the integral curvature $C(G)$ was defined in Sec. 3. For general G we proceed as follows: Let $G_1 \subset G_2 \subset \dots$ be a sequence of compact polygonal regions with $\bigcup G_i = G$ and the further property that a sequence of points $P_i \in G_{n_i+1} - G_{n_i}$, where $\{n_i\}$ is any increasing sequence of positive integers, has no accumulation point. If $\lim C(G_n)$ exists ($\pm \infty$ admitted) it is independent of the particular sequence $\{G_n\}$ and is called the *integral curvature* of G .

The condition that $\{P_i\}$ has no accumulation point implies for compact G that $G_n = G$ for large n , so that the present definition of $C(G)$ agrees with the previous one. It is necessary to add some such condition because $G = \bigcup G_i$ implies $C(G) = \lim C(G_i)$ in general only if $C(G)$ can be extended to a completely additive set function (compare Sec. 7).

If M is a G -surface of finite connectivity and the tubes T_i are defined as in the beginning of Sec. 5 and H is the compact domain on M bounded by the T_i , then $X(H)$ is independent of the choice of the T_i and $-X(H)$ is called the *characteristic* $-X(M)$ of M .

If $C(M)$ exists, it may be evaluated as follows: Let T_i^n be a sequence of sub tubes of T_i with $T_i^n \subset T_i^{n-1}$ and $\bigcap T_i^n = 0$. If H^n denotes the compact domain on M bounded by T_1^n, \dots, T_k^n then

$$C(M) = \lim C(H^n).$$

Since any tube, in particular T_i^n , contains a sub tube bounded by a polygon $R(u)$ as constructed in (16), it follows from (11) that

$$(27) \quad C(M) \leq 2\pi X(M) + k\pi,$$

provided $C(M)$ exists. The discussion preceding (22) yields

$$(28) \quad C(M) \geq 2\pi X(M) \text{ if } M \text{ has no expanding tubes.}$$

An application of (27) is

(29) *A non-compact surface with non-negative, but not identically vanishing curvature is homeomorphic to a plane.*

For there is a triangle on M with positive excess. This triangle contains then a triangle abc with positive excess which is so small that the images of abc on the universal covering surface M' of M are disjoint. If M were not a plane M' would have infinitely many sheets, and in each a copy of abc . The

integral curvature of M' , which exists because M' has non-negative curvature, is therefore ∞ . But this contradicts (27).

Finer results than (27) can be obtained if the angular metric on M is uniform:
(30) *If M has an integral curvature and a uniform angular metric then*

$$C(M) \leq 2\pi X(M).$$

The equality sign holds if M possesses no expanding tubes.

For if the previous notations are used, then T_ϵ^n carries a polygon $R_\epsilon^n(u_i)$ with the properties described in (23). If H^n is the compact domain bounded by the $R_\epsilon^n(u_i)$ then by (11)

$$C(H^n) \leq 2\pi X(M) + k'\epsilon.$$

The remark about the equality sign follows from (28).

Theorem (23) yields also

(31) *If the tube T with boundary P has an integral curvature and a uniform angular metric then*

$$C(T) \leq -\Sigma(\pi - \alpha_i),$$

where α_i are the angles of P measured in T .

If $C(T)$ exists, then every subtube of T has an integral curvature. If T is contracting or bulging, then it contains $T(u)$ bounded by an $R(u)$ which has one vertex q with a convex angle in T or is a closed geodesic. (31) yields then $C(T(u)) \leq 0$. Therefore

(32) *A tube with positive curvature and a uniform angular metric is expanding.*

Cohn-Vossen proves this for tubes with non-negative curvature, but the Riemannian character of his metric is essential for this refinement.

We next prove a theorem which is similar to (31) and is found in the paper [5] of Cohn-Vossen.

(33) *On M let Q be an open Jordan curve of the form $\bigcup_{i=-\infty}^{\infty} t(a_i, a_{i+1})$, $a_i \neq a_{i+1}$, and such that only a finite number of its angles are not straight. Assume, moreover, that Q bounds on M a domain G homeomorphic to a halfplane, and that each subarc of Q is a shortest connection of its endpoints in G . If $\alpha_1, \dots, \alpha_n$ are the angles at the proper vertices of Q in G and G has a uniform angular metric and an integral curvature, then*

$$C(G) \leq -\sum_{i=1}^n (\pi - \alpha_i).$$

Proof. Let G' be any simply connected domain in G bounded by a subarc Q' of Q which contains all n vertices of Q and a simple geodesic polygon Q'' in G connecting the two endpoints of Q' . Let $p \in Q'$ and let $q_1(t), q_2(t)$ be the two points of Q for which the subarcs from p to $q_i(t)$ of Q have length t . For a proper choice of t' the points $q_i(t)$ will lie on $Q - Q'$ for $t \geq t'$. Let $p(t)$ be a shortest connection of length $\lambda(t)$ of $q_1(t)$ and $q_2(t)$ in $(G - G') \cup Q''$. By the minimum property of Q

$$(34) \quad \lambda(t) \geq 2t.$$

As in the proof of (16) it is seen that $p(t)$ is a simple geodesic polygon whose

proper vertices, if any, coincide with vertices of Q'' and such that the corresponding angles are convex if measured in the domain $G(t)$ bounded by $p(t)$ and the subarc from $q_1(t)$ to $q_2(t)$ of Q . By (11)

$$C(G(t)) \leq 2\pi - (\pi - \beta_1(t) - (\pi - \beta_2(t)) - \Sigma(\pi - \alpha_i),$$

where $\beta_i(t)$ is the angle of $p(t)$ and Q at q_i measured in $G(t)$. Due to the arbitrariness of G' the theorem is proved if a $t_0 \geq t'$ exists for which $\beta_i(t_0) \leq \epsilon$. Let $k = 2\delta(\epsilon)$, where $\delta(\epsilon)$ is the function entering the definition of a uniform angular metric. Then because of (34)

$$(35) \quad \lambda(t) - 2t + kt \rightarrow \infty, \text{ for } t \rightarrow \infty.$$

The triangle inequality implies that $\lambda(t)$ is continuous, therefore the left side of (35) reaches a minimum at some value $t_0 \geq t'$. Then

$$(36) \quad \begin{aligned} \lambda(t_0 + h) - 2(t_0 + h) + k(t_0 + h) - \lambda(t_0) + 2t_0 - kt_0 &\geq 0, \text{ for } h > 0, \text{ or} \\ \lambda(t_0 + h) - \lambda(t_0) &\geq h(2 - k), \text{ for } h > 0. \end{aligned}$$

On the other hand if $a'_i, a''_i, a'_i \subset p(t_0), a''_i \subset Q(t_0)$, lie on the legs of the angle in $(G - G(t_0)) \cup p(t_0)$ at $q_i(t_0)$ (that means $|a'_i q_i(t_0) a''_i| = \pi - \beta_i(t_0)$) and satisfy the relations

$$a'_i q_i(t_0) = a''_i q_i(t_0) = h < \min_{1,2} \rho(q_i(t_0), \epsilon)$$

then $p(t_0 + h)$ is at most as long as the polygon originating from $p(t_0)$ by replacing $t(q_i(t_0), a'_i)$ by $t(a'_i, a''_i)$. Therefore

$$\lambda(t_0 + h) \leq \lambda(t_0) + (a'_1 a''_1 - h) + (a'_2 a''_2 - h)$$

which yields together with (36)

$$\begin{aligned} h(2 - k) &\leq a'_1 a''_1 - h + a'_2 a''_2 - h, \text{ or} \\ (1 - a'_1 a''_1 / 2h) + (1 - a'_2 a''_2 / 2h) &\leq k/2 = \delta(\epsilon). \end{aligned}$$

Since $1 - a'_i a''_i / 2h \geq 0$ it follows that $1 - a'_i a''_i / 2h \leq \delta(\epsilon)$ and from the definition of $\delta(\epsilon)$ that

$$\pi - \beta_i(t_0) = |a'_i q_i(t_0) a''_i| \geq \pi - \epsilon \quad \text{q.e.d.}$$

If in addition to the assumptions of (33) Q is a straight line (see [3, p. 232]) then there are no corners, hence $C(G) \leq 0$. Therefore

(37) *A plane with positive curvature does not contain a straight line.*

If the assumption that every subarc of Q is a shortest connection in G is omitted, Cohn-Vossen proves that

$$(38) \quad C(G) \leq \pi - \Sigma(\pi - \alpha_i).$$

In general spaces the inequality $C(G) \leq 2\pi - \Sigma(\pi - \alpha_i)$ is trivial, but the refinement from 2π to π rests on the fact that in Riemannian geometry perpendicular directions form the angle $\pi/2$. This fact has no analogue in general Finsler spaces, no matter how the angular metric is defined, because perpendicularity is not symmetric. Consequently there is no reason to believe that (38) holds with a suitable definition of angular measure, unless perpendicularity is symmetric, although the author did not try to construct an example because this would obviously be very laborious.

We conclude the analysis of the validity of Riemannian methods in general spaces, which could be continued almost *ad libitum*, by mentioning that the proofs of the following two interesting results of Cohn-Vossen [5] hold without any change:

Let M be a plane with positive curvature and uniform angular metric. Then every point of M lies on at least one geodesic without multiple points. If a geodesic g has multiple points, then it contains exactly one 1-gon P , moreover $g - P$ lies in the exterior of P and consists of two branches without multiple points (but the two branches may intersect each other).

7. The integral curvature as set function. On a surface M with a system S of geodesics and an angular metric as defined in Sec. 3, let F_0 be the collection of the following sets: the empty set, the points, the segments without end-points (1-cells), the interiors of the non-degenerate triangles (2-cells). The excess is called *completely additive* if for any representation of a 2-cell abc as union of a countable number of disjoint 2-cells a, b, c , and points and 1-cells on the boundaries of the a, b, c ,

$$\epsilon(abc) = \sum \epsilon(a, b, c).$$

The unions σ of a finite number of disjoint elements in F_0 form a field F_1 . If a, b, c , are the two cells of a given set $\sigma \in F_1$ we put

$$C(\sigma) = \sum \epsilon(a, b, c).$$

If $C(\sigma)$ is bounded on every bounded subset of M and the excess is completely additive, then $C(\sigma)$ can be extended to a completely additive set function on the σ -field F of all Borel sets on M . Moreover the extended set function is bounded on every bounded subset of M with the same bounds as the old function.⁹

Let a measure m be defined on M for which segments have measure 0, and such that every bounded measurable set on M has finite measure. We are going to prove the theorem

(39) *If for every bounded set B on M a number $\beta(B)$ exists such that for any 2-cell abc in B*

$$|\epsilon(abc)| \leq \beta(B)m(abc)$$

then ϵ is completely additive, $C(\sigma)$ is bounded on every bounded subset of M and absolutely continuous on F with respect to m .

Proof. Let σ be a set in F_1 which lies in the given set B and a, b, c , the 2-cells of σ . Then

$$C(\sigma) \leq \sum |\epsilon(a, b, c)| < \beta(B) \cdot \sum m(a, b, c) \leq \beta(B)m(\sigma) \leq \beta(B)m(B),$$

so that $C(\sigma)$ is bounded in B for $\sigma \in F_1$.

If abc is the union of the disjoint 2-cells a, b, c , $v = 1, 2, \dots$ and points and

⁹The arguments which lead to these conclusions are implicitly contained in many modern treatments of set functions. For those who are able to read Danish an unusually clear exposition is available in Jessen [7, part 3] which also determined the present formulation.

1-cells δ_i on the boundaries of the a, b, c , then $m(abc) = \sum m(a, b, c)$ because $m(\delta_i) = 0$. For a given $\epsilon > 0$ we can therefore find an $n(\epsilon)$ such that

$$m(abc) - \sum_{i=1}^{n(\epsilon)} m(a, b, c_i) < \epsilon/\beta(abc).$$

Then $abc - \sum_{i=1}^{n(\epsilon)} a, b, c_i$ is the sum of a finite number of 2-cells a'_i, b'_i, c'_i , $i = 1, \dots, m$ and a finite number of points and 1-cells. Therefore

$$\sum_{i=1}^m \epsilon(a'_i, b'_i, c'_i) \leq \beta(abc) \sum_{i=1}^m m(a'_i, b'_i, c'_i) = \beta(abc) m(abc - \sum_{i=1}^{n(\epsilon)} a, b, c_i) \leq \epsilon$$

which shows that $\epsilon(abc) = \sum_{i=1}^{\infty} \epsilon(a, b, c_i)$ or that ϵ is completely additive.

By the preceding remarks and the first part of this proof $C(\sigma)$, $\sigma \in F_1$, can be extended to a completely additive function on F with the same bounds. It then follows that

$$|C(\sigma)| \leq \beta(B)m(\sigma) \text{ for } \sigma \in F \text{ and } \sigma \subset B.$$

Therefore $m(\sigma) = 0$ implies $C(\sigma) = 0$ so that $C(\sigma)$ is absolutely continuous.

Under the hypotheses of the theorem, $C(\sigma)$ is therefore the indefinite integral $\int_{\sigma} f(p)$ of a function $f(p)$ with respect to the measure m . This does not yet assign a definite value to $f(p)$ at any given point since $f(p)$ can be changed at will in a set of measure 0. This indefiniteness can be eliminated if sufficient restrictions on the angular metric and on the measure guarantee that there is at least (and then exactly) one continuous $f(p)$. But it seems more worthwhile to discuss these questions in connection with a specific angular measure in a Finsler space.

REFERENCES

- [1] H. Busemann, "Metric Methods in Finsler Spaces and in the Foundations of Geometry," *Ann. Math. Studies* No. 8 (Princeton, 1942).
- [2] H. Busemann, "Local Metric Geometry," *Trans. Am. Math. Soc.*, vol. 56 (1944), 200-274.
- [3] H. Busemann, "Spaces with Non-positive Curvature," *Acta Math.*
- [4] S. Cohn-Vossen, "Kürzeste Wege und Totalkrümmung auf Flächen," *Comp. Math.*, vol. 2 (1935), 69-133.
- [5] S. Cohn-Vossen, "Totalkrümmung und geodätische Linien auf einfach zusammenhängenden, offenen, vollständigen Flächenstücken," *Mat. Sbornik*, N. S., vol. 1 (1936), 139-164.
- [6] J. Hadamard, "Les surfaces a courbures opposées et leur lignes géodésiques," *Jour. Math. Pur. Appl.*, 5th series, vol. 4 (1898), 27-73.
- [7] B. Jessen, *Abstrakt Maal-og Integralteori* (Copenhagen, 1947).
- [8] B.v.Kerékjártó, *Vorlesungen über Topologie I* (Berlin, 1923).
- [9] H. Seifert and W. Threlfall, *Lehrbuch der Topologie* (Leipzig, 1934).

University of Southern California

THE DENSITY OF REDUCIBLE INTEGERS

S. D. CHOWLA AND JOHN TODD

Introduction. The concept of a reducible integer was introduced recently [3] : if $P(m)$ denotes the greatest prime factor of m then n is said to be reducible if $P(1 + n^2) < 2n$. The reason for the term is that reducibility is a condition necessary and sufficient for the existence of a relation of the form

$$\arctan n = \sum_{i=1}^r f_i \arctan n_i$$

where the f_i are integers and the n_i positive integers less than n . J. C. P. Miller pointed out to us the regularity of the distribution of the reducible integers (less than 600). In collaboration with Dr. J. W. Wrench, using his tables of factors of $1 + n^2$, we carried the count still further, and observed the same regularity. The following conjecture suggested itself:

C. "Reducible integers have a density about 0.3."

We have not been able to make very much headway with this but have succeeded in establishing the following:

THEOREM A. *The density of the set of integers n for which $P(n) < 2n^{\frac{1}{2}}$ is $1 - \log 2 = .3069 \dots$*

This note contains a proof of this theorem, and a table summarizing the numerical evidence in support of C.

1. Numerical evidence. We give here a summary of the numerical evidence relating to the conjecture C together with corresponding results related to Theorem A. The table below gives, in each range $(1 + 100n, 100(n + 1))$, for $n = 0(1)49$, on the right, the number of reducible integers in that range, and on the left, the number of integers in that range which satisfy $P(n) < 2n^{\frac{1}{2}}$.

Totals in the various chiliads and a grand total for the complete range (1-5000) are given in the last line of the table.

	0	1000	2000	3000	4000
1-100	(29, 57)	(31, 43)	(29, 43)	(33, 41)	(29, 42)
101-200	(29, 50)	(25, 43)	(30, 42)	(28, 43)	(28, 40)
201-300	(28, 47)	(33, 44)	(23, 42)	(23, 43)	(27, 41)
301-400	(26, 45)	(28, 41)	(32, 41)	(32, 43)	(31, 40)
401-500	(30, 45)	(31, 44)	(28, 44)	(29, 38)	(27, 42)
501-600	(30, 44)	(23, 44)	(32, 39)	(32, 41)	(38, 39)
601-700	(30, 44)	(27, 40)	(26, 43)	(25, 40)	(30, 41)
701-800	(29, 44)	(34, 43)	(32, 41)	(30, 43)	(35, 39)
801-900	(27, 44)	(28, 45)	(27, 42)	(29, 40)	(30, 43)
901-1000	(23, 42)	(31, 39)	(29, 41)	(19, 41)	(38, 41)
	(281, 462)	(291, 426)	(288, 418)	(280, 413)	(313, 408)
	(1453, 2127)				

Received April 5, 1948. This work has been supported by the Office of Naval Research of the United States Navy Department.

2. Proof of Theorem A. It is more convenient to show that the density of the integers n for which $P(n) \geq 2n^{\frac{1}{2}}$ is $\log 2$. That is, we shall show that

$$Q(x) = \sum_{\substack{n \leq x \\ P(n) \geq 2n^{\frac{1}{2}}}} 1 \sim x \log 2;$$

to do this we establish the two following results:

$$A_1. \quad Q_1(x) = \sum_{\substack{n \leq x \\ P(n) \geq 2x^{\frac{1}{2}}}} 1 \sim x \log 2;$$

$$A_2. \quad Q_2(x) = Q(x) - Q_1(x) = \sum_{\substack{n \leq x \\ 2n^{\frac{1}{2}} \leq P(n) \leq 2x^{\frac{1}{2}}}} 1 = o(x).$$

2.1. Proof of A_1 . This is carried out by a modification of a method used recently [1] to evaluate $\lim x^{-1}R_n(x)$ where $R_n(x)$ is the number of integers $n \leq x$ for which $P(n) \geq x^{\frac{1}{2}}$.

For any p the number of integers $n \leq x$ which are multiples of p is $[x/p]$. In $Q_1(x)$ we consider only primes $p = P(n) \geq 2x^{\frac{1}{2}}$: for such primes the residual factor $(n/p) \leq \frac{1}{2}x^{\frac{1}{2}} < p$ and so every multiple of p which does not exceed x has p for its greatest prime factor. Hence

$$\begin{aligned} Q_1(x) &= \sum_{2x^{\frac{1}{2}} \leq p \leq x} [x/p] \\ &= \sum_{2x^{\frac{1}{2}} \leq p \leq x} \{(x/p) + O(1)\} \\ &= x \sum_{2x^{\frac{1}{2}} \leq p \leq x} p^{-1} + O(x/\log x), \end{aligned}$$

since $\sum_{2x^{\frac{1}{2}} \leq p \leq x} 1 \leq \sum_{p \leq x} 1 = O(x/\log x)$.

It is, however, well known [2, pp. 100-102] that

$$B. \quad \sum_{p \leq x} p^{-1} = \log \log x - l + O(1/\log x)$$

where l is a certain constant. Hence

$$\begin{aligned} x^{-1}Q_1(x) &= \log \log x - \log \log 2x^{\frac{1}{2}} + o(1) \\ &= \log \{(\log x)/(\frac{1}{2} \log x + \log 2)\} + o(1) \\ &= \log 2 + o(1), \end{aligned}$$

which establishes A_1 .

2.2. Proof of A_2 . This is carried out in the following manner. First, it will be sufficient to restrict the values of n considered to the range

$$x/(\log x)^2 \leq n \leq x,$$

for this implies a change in the sum of $O(x/(\log x)^2) = o(x)$. Secondly, we do not decrease the sum if we replace $2n^{\frac{1}{2}}$, the variable limit in the lower inequality, by its smallest value $2x^{\frac{1}{2}}/\log x$. Thirdly, we do not decrease the sum by now allowing n to cover the full range $1 \leq n \leq x$. Thus it will be sufficient to show that

$$Q_3(x) = \sum_{\substack{n \leq x \\ (2x^{\frac{1}{2}}/\log x) \leq p(n) \leq 2x^{\frac{1}{2}}}} 1 = o(x).$$

In order that an integer should contribute to Q_2 it is necessary that it should have a prime factor p in the range $(2x^{1/2}/\log x, 2x^{1/2})$. For p fixed the number of such n is $[x/p]$. Hence

$$Q_2 \leq \sum_{(2x^{1/2}/\log x) \leq p \leq 2x^{1/2}} [x/p].$$

(It is possible for an integer $n \leq x$ to have two factors in the range and so we must allow for inequality, which was not so in the case of Q_1 .)

We now proceed as before:

$$\begin{aligned} Q_2(x) &\leq \sum_{(2x^{1/2}/\log x) \leq p \leq 2x^{1/2}} [x/p] = x \sum_{(2x^{1/2}/\log x) \leq p \leq 2x^{1/2}} p^{-1} + O(x^{1/2}/\log x) \\ &= x \{ \log \log 2x^{1/2} - \log \log (2x^{1/2}/\log x) \} + O(x/\log x), \end{aligned}$$

using B. Since

$$\begin{aligned} &\log \log 2x^{1/2} - \log \log (2x^{1/2}/\log x) \\ &= \log \{ (\tfrac{1}{2} \log x + \log 2) / (\tfrac{1}{2} \log x + \log 2 - \log \log x) \} \\ &= \log \{ \{ 1 + (\log 4)/(\log x) \} \{ 1 + (\log 4 - 2 \log \log x)/\log x \}^{-1} \} \\ &= \log \{ 1 + O(1/\log x) \} (1 + O(\log \log x / \log x)) \\ &= O(\log \log x / \log x) = o(1), \end{aligned}$$

the proof of A_2 is complete.

3. Possible generalizations. It is clear that $2n^{1/2}$ in Theorem A can be replaced by $An^{1/2}$ for any $A \geq 1$ without affecting the conclusion.

Similar arguments show that the density of the integers n for which $P(n) > An^a$ ($\frac{1}{2} < a < 1$, $A > 1$) is exactly $\log a$.

The case when $a < \frac{1}{2}$ requires more careful study along the lines indicated in [1] and it can be shown that the device used here (replacing a summation over $1 \leq n \leq x$ by one over $(x/(\log x)^2 \leq n \leq x)$) will enable the density to be evaluated explicitly in this case, too.

It is clear that an estimate for the error term

$$x^{-1}Q(x) - \log 2$$

is

$$O(\log \log x / \log x),$$

and this explains the slowness of the convergence apparent in the table.

REFERENCES

- [1] S. Chowla and T. Vijayaraghavan, *J. Indian Math. Soc. (New Series)*, vol. 11 (1947), 31-37.
- [2] E. Landau, *Handbuch der Lehre von der Verteilung der Primzahlen* (Leipzig, 1909).
- [3] John Todd, "A Problem of J. C. P. Miller on Arctangent Relations," *Amer. Math. Monthly* (1949).

*Institute for Advanced Study
King's College, London*

ON A THEOREM OF LATIMER AND MACDUFFEE

OLGA TAUSKY

THE matrix solutions of an irreducible algebraic equation with integral coefficients were studied by Latimer and MacDuffee.¹ They considered matrices with rational integers as elements. If A is such a matrix, then all matrices of the "class" $S^{-1}AS$ will again be solutions if S is a matrix of determinant ± 1 . On the other hand, in general all solutions cannot be derived in this way from one solution only. It was in fact shown that the number of classes of matrix solutions coincides with the number of different classes of ideals in the ring generated by an algebraic root of the same equation. Although this result is of interest in many different branches of mathematics it is not generally known. It seems particularly often required for periodic matrices.²

Latimer and MacDuffee actually dealt with the more general case when the equation was reducible. By restriction to irreducible equations only, a very simple proof can be obtained.

In what follows $f(x)=0$ is an irreducible algebraic equation of degree n with integral coefficients, α one of its algebraic roots, $A=(a_{ik})$ an $n \times n$ matrix with rational integers as elements which satisfies $f(x)=0$ and S is a matrix with rational integers as elements and determinant ± 1 .

THEOREM 1. *The algebraic number α is a characteristic root of the matrix A and the components of the corresponding characteristic vector (a_1, \dots, a_n) can be chosen to form the basis of an ideal in the ring formed by the polynomials in α with rational integers as coefficients.*

Proof. Since $f(x)$ is assumed irreducible it follows that it is the characteristic and the minimum polynomial of A and that α is a characteristic root of A . Since in this case the characteristic roots of A are all simple, the corresponding characteristic vector is uniquely determined apart from a factor of proportionality. Since

$$(1) \quad \alpha(a_1, \dots, a_n) = A(a_1, \dots, a_n)$$

we may take for a_i the cofactor of the i th element in a fixed row of the deter-

Received July 20, 1948.

¹C. G. Latimer and C. C. MacDuffee, "A Correspondence Between Classes of Ideals and Classes of Matrices," *Ann. of Math.*, vol. 34 (1933), 313-316. See also related work in A. Speiser, *Theorie der Gruppen* (Springer, 1937); B. L. van der Waerden, *Gruppen von linearen Transformationen* (Springer, 1935); H. Zassenhaus, "Neuer Beweis der Endlichkeit der Klassenzahl bei unimodularer Äquivalenz endlicher ganzzahliger Substitutionsgruppen," *Abh. Math. Sem. Hansischen Univ.*, vol. 12 (1938), 276-288.

²See e.g. R. P. Bambah and S. Chowla, "On Integer Roots of the Unit Matrix," *Proc. Nat. Inst. Sci. India*, vol. 13 (1937), 241-246.

minant $|a_{ik} - a\delta_{ik}|$. This is a polynomial in a with rational integral coefficients. From (1) it follows that

$$a^i(a_1, \dots, a_n) = A^i(a_1, \dots, a_n), \quad (i = 0, \dots, n-1).$$

Since the numbers $1, a, \dots, a^{n-1}$ form a basis for the ring in question, it is proved that the set of numbers

$$a_1a_1 + \dots + a_na_n$$

where a_i are rational integers, forms an ideal.

THEOREM 2. *Two ideals determined (as in Theorem 1) from the same matrix A belong to the same ideal class.*

Proof. Since the elements of the basis of an ideal in Theorem 1 are uniquely determined apart from a common multiplier it follows that any two such ideals must be equivalent—as usual two ideals a and b are said to be equivalent or belong to the same class if two elements, α, β in the ring exist such that

$$a\alpha = b\beta.$$

THEOREM 3. *To every ideal $(\omega_1, \dots, \omega_n)$ in the ring generated by a there corresponds a matrix X with rational integers as elements which satisfies $f(X) = 0$ and such that*

$$a(\omega_1, \dots, \omega_n) = X(\omega_1, \dots, \omega_n).$$

Proof. Since $(\omega_1, \dots, \omega_n)$ is an ideal there must exist a relation

$$(2) \quad a(\omega_1, \dots, \omega_n) = X(\omega_1, \dots, \omega_n)$$

where X is a matrix with rational integral elements. From (2) follows

$$a^i(\omega_1, \dots, \omega_n) = X^i(\omega_1, \dots, \omega_n), \quad (i = 0, \dots, n-1).$$

This implies

$$f(a)(\omega_1, \dots, \omega_n) = f(X)(\omega_1, \dots, \omega_n) = 0.$$

Since $f(X)$ is also a matrix with rational elements and since the relations

$$f(X)(\omega_1^{(0)}, \dots, \omega_n^{(0)}) = 0$$

hold in the fields generated by the conjugate roots of a and $|\omega_i^{(k)}| \neq 0$ it follows that

$$f(X) = 0.$$

THEOREM 4. *The matrix X in Theorem 3 is uniquely determined apart from a transformation SXS^{-1} .*

Proof. If a different basis for the ideal were chosen it would be of the form $S(\omega_1, \dots, \omega_n)$ with $|S| = \pm 1$. There would then be a relation

$$aS(\omega_1, \dots, \omega_n) = YS(\omega_1, \dots, \omega_n).$$

On the other hand, in virtue of (2)

$$aS(\omega_1, \dots, \omega_n) = SX(\omega_1, \dots, \omega_n).$$

Hence by the argument used at the end of the proof of Theorem 3:

$$\begin{aligned} SX &= YS & \text{or} \\ Y &= SXS^{-1}. \end{aligned}$$

The Theorems 1-4 show that *there is a 1-1 correspondence between the classes of matrices and the ideal classes.*

It may be pointed out that the matrices S for which $SAS^{-1} = A$ play a role similar to the units in the algebraic number fields. Such a matrix is in fact a polynomial⁸ in A , and since its determinant is ± 1 it is a unit in the field generated by A .

*Institute for Numerical Analysis
National Bureau of Standards*

⁸See e.g. J. H. M. Wedderburn, "Lectures on Matrices," *Amer. Math. Soc. Colloquium Publications*, vol. 17 (1934), 27.

CONGRUENCE RELATIONS BETWEEN THE TRACES OF MATRIX POWERS

J. S. FRAME

1. Introduction. Let A be a matrix of finite order n and finite degree d , whose characteristic roots are certain n^{th} roots of unity $\alpha_1, \alpha_2, \dots, \alpha_d$. We wish to prove a congruence (6) between the traces (tr) of certain powers of A , which is suggested by two somewhat simpler congruences (1) and (3).

First, if $\text{tr}(A)$ is a rational integer, it is easy to establish the familiar congruence

$$(1) \quad \text{tr}(A) \equiv \text{tr}(A^p) \pmod{p}, \quad p \text{ prime,}$$

even though $\text{tr}(A^p)$ may not itself be rational. For we have

$$(2) \quad [\text{tr}(A)]^p = \left[\sum_{r=1}^d \alpha_r \right]^p = \sum_{r=1}^d \alpha_r^p + p(\dots) = \text{tr}(A^p) + p(\dots)$$

where (\dots) denotes an algebraic integer. The left-hand members of (1) and (2) are rational integers which are congruent \pmod{p} by Fermat's theorem. The right-hand members are explicitly congruent \pmod{p} . Hence (1) follows from (2).

Secondly, for any integer a , we have

$$(3) \quad a^{p^\theta} \equiv a^{p^\theta-1} \pmod{p^\theta}, \text{ if } p^\theta \text{ is a prime power } > 1.$$

Equation (3) is trivial if a is divisible by p . Otherwise it can be established easily by setting $m = p^\theta$ in the well-known Euler congruence

$$(4) \quad a^{\phi(m)} \equiv 1 \pmod{m}, \quad \text{for } (a, m) = 1,$$

where $\phi(m)$ is the Euler ϕ -function, and $\phi(p^\theta) = p^\theta - p^{\theta-1}$.

It is our purpose to prove a congruence relation $\pmod{p^\theta}$, which generalizes (1) and is similar to (3), between the traces of certain powers of a matrix A of finite order—or, in other words, between certain sums of powers of roots of unity.

THEOREM. Let $S(m)$ denote the trace of the m^{th} power of a matrix A of finite order n and finite degree $S(0)$, and assume that A is such that

$$(5) \quad S(k) = S(1), \text{ for all } k \text{ such that } (k, n) = 1.$$

Then

$$(6) \quad S(p^\theta) \equiv S(p^{\theta-1}) \pmod{p^\theta}.$$

We note that condition (5) implies that A has a rational integral trace, but that not every matrix with rational integral trace satisfies (5).

Received July 21, 1948.

2. Proof of the theorem.¹ Let us define a " p^θ -set" to be a set of roots of unity such that the sum of its $p^{\theta h}$ powers are congruent to the sum of its $p^{\theta-1h}$ powers mod p^θ as in (6). We note that the negative of any root of unity is also a root of unity.

LEMMA 1. *The set of all the n distinct n^{th} roots of unity is a p^θ -set.*

Proof. Denoting the sum of m^{th} powers by $S_n(m)$ we have

$$(7) \quad S_n(p^\theta) = \begin{cases} n, & \text{if } n \text{ divides } p^\theta, \\ 0, & \text{if } n \text{ does not divide } p^\theta. \end{cases}$$

Hence

$$(8) \quad S_n(p^\theta) - S_n(p^{\theta-1}) = \begin{cases} p^\theta & \text{if } n = p^\theta, \\ 0 & \text{otherwise.} \end{cases}$$

LEMMA 2. If one p^θ -set is included as a subset of a larger p^θ -set, the difference of the two p^θ -sets is also a p^θ -set. Furthermore, any set of roots of unity which is made up of two or more p^θ -sets is also a p^θ -set.

Proof. If each of two or more quantities $S(p^\theta) - S(p^{\theta-1})$ is congruent to 0, so is their sum or difference.

LEMMA 3. The set of $\phi(n)$ primitive n^{th} roots of unity is a p^θ -set.

For prime n the lemma is a special case of Lemma 2. Assuming as induction hypothesis that the lemma is true for all ν with a smaller number of prime factors than n , we show that it is also true for n by applying Lemmas 1 and 2, and eliminating from the complete set of n n^{th} roots all sets of primitive ν^{th} roots for each ν which is a proper divisor of n . Only the primitive n^{th} roots remain. They form a p^θ -set.

We observe that condition (5) implies that for any factor μ of n the primitive μ^{th} roots occur as roots of the matrix A with equal multiplicity. Hence by Lemmas 2 and 3 the roots of A are a p^θ -set, so the theorem is established.

3. Applications of the theorem. In constructing the table of characters for a finite group, our theorem may be applied to determine many of the entries. For example, the symmetric group of degree 5 and order 120 has irreducible representations of degrees 1, 1, 4, 4, 5, 5, 6. There are 15 conjugate elements of order 2 which are squares of elements of order 4. Hence their traces form a vector of unitary squared length 120/15 which is unitary orthogonal to the vector (1, 1, 4, 4, 5, 5, 6) and congruent to it (mod 4). The only integral solution is (1, 1, 0, 0, 1, 1, -2). Similarly for the traces of the 24 elements of order 5 we have the vector (1, 1, -1, -1, 0, 0, 1) as is known by the ordinary modular theory (mod 5). Given the numbers of elements in the classes of conjugates, the table is completely determined by these congruences.

Michigan State College

¹I am indebted to Professor R. Brauer for some suggestions for shortening my original proof.

